GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 16:39:16 ; Search time 32 Seconds
(without alignments)
715.038 Million cell updates/sec

Title: US-09-674-266A-181
Perfect score: 1063
Sequence: 1 RLSCAGTLSGSGPHPSRRLT.....DTALLLDNMKKALKLLKTEL 206

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1063	100.0	206	20	AAY73837	Human prostate tum
2	1002	94.3	194	20	AAY76590	Human ovarian tumo
3	1002	94.3	202	22	AAG75621	Human colon cancer
4	899	84.6	175	19	AAW77365	zsig10 polypeptide
5	899	84.6	175	19	AAW37844	Human XAG growth f
6	899	84.6	175	19	AAW37872	Human protein comp
7	899	84.6	175	20	AAY59675	Secreted protein 1
8	899	84.6	175	21	AAB24070	Human PRO1030 prot
9	899	84.6	175	22	AAU08804	Breast cancer cell
10	899	84.6	175	22	AAB72203	Human huXAG-1/CCSG
11	656	61.7	132	21	AAG00109	Human secreted pro
12	605	56.9	131	20	AAY59718	Secreted protein 7
13	598	56.3	116	20	AAY12312	Human 5' EST secre
14	593	55.8	115	21	AAY64672	Human 5' EST relat
15	555	52.2	166	19	AAW37846	Human XAG growth f
16	555	52.2	166	21	AAB00194	Breast cancer prot
17	555	52.2	166	22	AAU25728	Breast cancer-asso
18	555	52.2	166	22	AAB72205	Human huXAG-3/CCSG
19	555	52.2	166	22	AAB31192	Amino acid sequenc
20	555	52.2	166	22	AAU07647	Human BCMP 11 poly
21	555	52.2	168	22	AAM24502	Colon tumour relat
22	388	36.5	89	20	AAY76568	Human ovarian tumo
23	361	34.0	180	22	AAG73737	Human colon cancer
24	285	26.8	56	20	AAY11881	Human 5' EST secre
25	222	20.9	56	20	AAY11939	Human 5' EST secre
26	222	20.9	172	19	AAW37845	Human XAG growth f
27	222	20.9	172	20	AAY03236	Clone HP01766 of a
28	222	20.9	172	20	AAY03237	Clone HP01766 of a
29	222	20.9	172	20	AAY03230	Amino acid sequenc
30	222	20.9	172	21	AAB10282	Xenopus sp embryo
31	222	20.9	172	21	AAB24432	Human PRO1376 prot
32	222	20.9	172	22	AAB88366	Human membrane or
33	222	20.9	172	22	AAB72204	Human huXAG-2/CCSG
34	218.5	20.6	172	19	AAW80812	Amino acid sequenc
35	210	19.8	62	20	AAY11882	Human 5' EST secre
36	179.5	16.9	70	20	AAY11654	Human 5' EST secre
37	88.5	8.3	118	20	AAY35816	Amino acid sequenc
38	88.5	8.3	1788	22	ABG06749	Novel human diagno
39	87.5	8.2	458	22	AAG70789	S cerevisiae apopt
40	86.5	8.1	1614	22	ABG17640	Novel human diagno
41	84	7.9	2027	22	ABG07898	Novel human diagno
42	83.5	7.9	1409	22	ABB65706	Drosophila melanog
43	83	7.8	1146	22	ABG14251	Novel human diagno
44	82	7.7	556	22	ABG13300	Novel human diagno
45	82	7.7	595	22	ABG11877	Novel human diagno

ALIGNMENTS

RESULT 1
AAY73837
ID AAY73837 standard; Protein; 206 AA.
XX
AC AAY73837;
XX
DT 14-MAR-2000 (first entry)
XX
DE Human prostate tumor EST fragment derived protein #24.
XX
KW Pancreas; tumor; EST; expressed sequence tag; human; cytostatic;
KW treatment.
XX
OS Homo sapiens.
XX
PN DE19820190-A1.
XX
PD 04-NOV-1999.
XX
PF 28-APR-1998; 98DE-1020190.
XX
PR 28-APR-1998; 98DE-1020190.
XX
PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
XX
DR WPI; 1999-621386/54.
DR N-PSDB; AAZ52865.
XX
PT New human nucleic acid sequences from pancreatic tumors, and related
PT proteins -
XX
PS Claim 23; Page 318; 502pp; German.

```
XX
CC   This invention describes novel polypeptides and their encoding nucleic
CC   acids derived from human pancreatic tumor tissue which have cytostatic
CC   activity. The sequences are also useful in producing pharmaceutical
CC   compositions for treatment of pancreatic tumors. AAY73814-Y74252
CC   represent protein fragments encoded by the human pancreatic tumor cDNA
CC   library derived expressed sequence tag (EST) sequences represented in
CC   AAZ52858-Z53014.
XX
SQ   Sequence   206 AA;

  Query Match           100.0%;  Score 1063;  DB 20;  Length 206;
  Best Local Similarity  100.0%;  Pred. No. 2.3e-108;
  Matches  206;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 RLSCAGTLSGSGPHPSRRLTQGRWVRKSRVAMEKIPVSAFLLLVALSYTLARDTTVKPGA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 RLSCAGTLSGSGPHPSRRLTQGRWVRKSRVAMEKIPVSAFLLLVALSYTLARDTTVKPGA 60

Qy      61 KKDTKDSRPKLPQTLSRGWGDQLIWTQTYEEALYKSKTSNKPLMIIHHLDECPHSQALKK 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 KKDTKDSRPKLPQTLSRGWGDQLIWTQTYEEALYKSKTSNKPLMIIHHLDECPHSQALKK 120

Qy     121 VFAENKEIQKLAEQFVLLNLVYETTDKHLSPDGQYVPRIMFVDPSLTVRADITGRYSNRL 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 VFAENKEIQKLAEQFVLLNLVYETTDKHLSPDGQYVPRIMFVDPSLTVRADITGRYSNRL 180

Qy     181 YAYEPADTALLLDNMKKALKLLKTEL 206
          |||||||||||||||||||||||||||
Db     181 YAYEPADTALLLDNMKKALKLLKTEL 206


RESULT 2
AAY76590
ID   AAY76590 standard; Protein; 194 AA.
XX
AC   AAY76590;
XX
DT   10-APR-2000  (first entry)
XX
DE   Human ovarian tumor EST fragment encoded protein 86.
XX
KW   Expressed sequence tag; EST; human; ovarian tumor; anticancer;
KW   gene therapy; treatment.
XX
OS   Homo sapiens.
XX
PN   DE19817557-A1.
XX
PD   21-OCT-1999.
XX
PF   09-APR-1998;   98DE-1017557.
XX
PR   09-APR-1998;   98DE-1017557.
XX
PA   (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI   Rosenthal A,  Specht T,  Hinzmann B,  Schmitt A,  Pilarsky C,  Dahl E;
XX
DR   WPI; 1999-591920/51.
DR   N-PSDB; AAZ77486.
XX
PT   New nucleic acid sequences expressed in ovarian, and some other, cancer
PT   tissues, and derived polypeptides, for treatment of ovarian cancer and
PT   identification of therapeutic agents -
XX
PS   Claim 25; Page 279; 310pp; German.
XX
CC   This invention describes novel nucleic acid (cDNA) sequences (A) which
CC   have anticancer activity and are highly expressed in ovarian tumor
CC   tissue (and some also in testis and breast cancer tissue). The products
CC   of the invention can be used for gene therapy. (A) are used (i) for
CC   recombinant expression of polypeptides (B) and (ii) to isolate complete
CC   genes. (B) are used (i) to identify agents suitable for treatment of
CC   ovarian cancer; (ii) directly for treating this form of cancer
CC   (including expression from gene therapy vectors) and (iii) for generation
CC   of specific antibodies. (A) are identified by assembling ESTs (expressed
CC   sequence tags) from a particular tissue type before comparison of
CC   expression patterns. This allows a significantly longer fragment of the
CC   gene to be revealed, so should reduce the number of failures associated
CC   with the fact that ESTs from different libraries may represent different
CC   parts of the same unknown gene, distorting the estimated frequency of
CC   occurrence in a particular tissue. AAY76505-Y76638 represent protein
CC   fragments encoded by the human ovarian tumor cDNA library derived EST
CC   fragments represented in AAZ77450-Z77572.
XX
SQ   Sequence   194 AA;

  Query Match           94.3%;  Score 1002;  DB 20;  Length 194;
  Best Local Similarity  100.0%;  Pred. No. 1e-101;
  Matches  194;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      13 PHPSRRLTQGRWVRKSRVAMEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLP 72
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 PHPSRRLTQGRWVRKSRVAMEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLP 60

Qy      73 QTLSRGWGDQLIWTQTYEEALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLA 132
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 QTLSRGWGDQLIWTQTYEEALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLA 120

Qy     133 EQFVLLNLVYETTDKHLSPDGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLL 192
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 EQFVLLNLVYETTDKHLSPDGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLL 180

Qy     193 DNMKKALKLLKTEL 206
           ||||||||||||||
Db     181 DNMKKALKLLKTEL 194


RESULT 3
AAG75621
ID   AAG75621 standard; Protein; 202 AA.
XX
AC   AAG75621;
XX
DT   03-SEP-2001  (first entry)
XX
DE   Human colon cancer antigen protein SEQ ID NO:6385.
XX
KW   Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW   colorectal carcinoma.
XX
OS   Homo sapiens.
XX
PN   WO200122920-A2.
XX
PD   05-APR-2001.
XX
PF   28-SEP-2000; 2000WO-US26524.
XX
PR   29-SEP-1999;   99US-0157137.
PR   03-NOV-1999;   99US-0163280.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Barash SC,  Birse CE,  Rosen CA;
XX
DR   WPI; 2001-235357/24.
DR   N-PSDB; AAH35026.
XX
PT   Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT   useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS   Claim 11; Page 7865-7866; 9803pp; English.
XX
CC   AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
```

```
CC   cancer-associated nucleic acid molecules (N) and proteins (P), where
CC   the proteins are collectively known as colon cancer antigens. The colon
CC   cancer antigens have cytostatic activity and can be used in gene
CC   therapy and vaccine production. N and P may be used in the prevention,
CC   diagnosis and treatment of diseases associated with inappropriate P
CC   expression. For example, N and P may be used to treat disorders
CC   associated with decreased expression by rectifying mutations or deletions
CC   in a patient's genome that affect the activity of P by expressing
CC   inactive proteins or to supplement the patients own production of P.
CC   Additionally, N may be used to produce the colon cancer-associated Ps,
CC   by inserting the nucleic acids into a host cell and culturing the cell
CC   to express the proteins. N and P can be used in the prevention, diagnosis
CC   and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC   and AAB77789 represent sequences used in the exemplification of the
CC   present invention.
CC   N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC   missing at time of publication, meaning no sequences are present for
CC   SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ   Sequence   202 AA;

  Query Match             94.3%;  Score 1002;  DB 22;  Length 202;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-101;
  Matches  194;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      13 PHPSRRLTQGRWVRKSRVAMEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLP 72
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9 PHPSRRLTQGRWVRKSRVAMEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLP 68

Qy      73 QTLSRGWGDQLIWTQTYEEALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLA 132
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69 QTLSRGWGDQLIWTQTYEEALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLA 128

Qy     133 EQFVLLNLVYETTDKHLSPDGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLL 192
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     129 EQFVLLNLVYETTDKHLSPDGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLL 188

Qy     193 DNMKKALKLLKTEL 206
           ||||||||||||||
Db     189 DNMKKALKLLKTEL 202


RESULT 4
AAW77365
ID   AAW77365 standard; Protein; 175 AA.
XX
AC   AAW77365;
XX
DT   21-DEC-1998  (first entry)
XX
DE   zsig10 polypeptide.
XX
KW   Human; mucous-mediated function; adhesion; tumour metastasis;
KW   bacterial colonisation; microbial infection; AIDS; cystic fibrosis;
KW   chronic obstructive pulmonary disease; asthma; Crohn's disease;
KW   sinonasal inflammatory disease; inflammatory bowel disease; bronchitis.
XX
OS   Homo sapiens.
XX
PN   WO9841627-A1.
XX
PD   24-SEP-1998.
XX
PF   18-MAR-1998;   98WO-US05251.
XX
PR   19-MAR-1997;   97US-0039631.
XX
PA   (ZYMO ) ZYMOGENETICS INC.
XX
PI   Sheppard PO;
XX
DR   WPI; 1998-531566/45.
DR   N-PSDB; AAV59320.
XX
PT   New isolated mucous-associated polypeptide, zsig10 - used to develop
PT   products for treating e.g. tumour metastasis, microbial infections,
PT   cystic fibrosis, asthma, bronchitis or inflammatory bowel disease
XX
PS   Claim 1; Page 82; 109pp; English.
XX
CC   The human polypeptide zsig10 is involved in mucous-mediated functions
CC   such as adhesion.  The products of the invention can be used in the study
CC   and treatment of e.g. tumour metastasis, bacterial colonisation,
CC   susceptibility to and persistence of infection, microbial infections,
CC   AIDS, cystic fibrosis, chronic obstructive pulmonary disease, asthma,
CC   sinonasal inflammatory disease, inflammatory bowel disease, bronchitis,
CC   or Crohn's disease.  The products can also be used for detection,
CC   diagnosis and drug screening.
XX
SQ   Sequence   175 AA;

  Query Match             84.6%;  Score 899;  DB 19;  Length 175;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-90;
  Matches  175;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      32 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 91
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 60

Qy      92 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 151
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 120

Qy     152 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 206
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 175


RESULT 5
AAW37844
ID   AAW37844 standard; Protein; 175 AA.
XX
AC   AAW37844;
XX
DT   28-AUG-1998  (first entry)
XX
DE   Human XAG growth factor huXAG-1.
XX
KW   HuXAG-1; XAG; growth factor; colon cancer; tumour marker;
KW   breast disease; liver disease; lung disease; emphysema;
KW   wound healing; diagnosis; therapy; human.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..20
FT                   /label= Sig_peptide
FT   Protein         21..175
FT                   /label= Mat_protein
FT                   /note= "Claim 11"
FT   Peptide         20..43
FT                   /label= Epitope
FT                   /note= "Claim 12"
FT   Peptide         44..52
FT                   /label= Epitope
FT                   /note= "Claim 12"
FT   Peptide         61..72
FT                   /label= Epitope
FT                   /note= "Claim 12"
FT   Peptide         90..103
FT                   /label= Epitope
FT                   /note= "Claim 12"
FT   Peptide         113..125
FT                   /label= Epitope
FT                   /note= "Claim 12"
FT   Peptide         138..150
```

FT /label= Epitope
FT /note= "Claim 12"
XX
PN WO9807749-A1.
XX
PD 26-FEB-1998.
XX
PF 22-AUG-1997; 97WO-US14139.
XX
PR 23-AUG-1996; 96WO-US13766.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Dillon PJ, Ebner R, Endress GA, Yu G;
XX
DR WPI; 1998-169093/15.
DR N-PSDB; AAV19155.
XX
PT New isolated human XAG growth factor(s) - used to develop products
PT for treating e.g. liver, lung or breast diseases or
PT hyperproliferative disorders, e.g. cancer.
XX
PS Claim 1; Fig 1; 141pp; English.
XX
CC This polypeptide comprises huXAG-1, a member of a novel family of
CC human growth factors also including huXAG-2 (see AAW37845) and
CC huXAG-3 (see AAW37846). These proteins share homology with the
CC XAG protein of Xenopus laevis, which is involved in embryogenesis
CC and is expressed in adult tissue. huXAG-1 is specifically found
CC in cancerous colon cells and may therefore be a growth factor for
CC colon cancer. huXAG-1 cDNA (see AAV19155) was isolated from a cDNA
CC library derived from human colon cancer tissue. Vectors, host
CC cells, antibodies, and screening methods for identifying agonists
CC and antagonists of huXAG-1 are provided. HuXAG polypeptides are
CC growth factors and can be used to stimulate proliferation of cells.
CC They can be used to stimulate the proliferation and differentiation
CC of hepatocytes to alleviate or treat liver diseases and pathologies
CC such as fulminant liver failure caused by cirrhosis, liver damage
CC caused by viral hepatitis and toxic substances. They can also be
CC used to stimulate or promote liver regeneration, e.g. after
CC surgery. They can also be used to prevent and heal damage to the
CC lungs caused by various pathological states. They can be used to
CC stimulate proliferation and differentiation and promote the repair
CC of alveoli and bronchiolar epithelium to prevent, attenuate, or
CC treat acute or chronic lung damage, e.g. emphysema, which results
CC in the progressive loss of alveoli, and inhalation injuries, e.g.
CC resulting from smoke inhalation and burns, that cause necrosis of
CC the bronchiolar epithelium and alveoli. They can also be used to
CC stimulate the proliferation and differentiation of breast tissue
CC and could therefore be used to promote healing of breast tissue
CC injury due to surgery, trauma or cancer. Antagonists can be used
CC to treat hyperproliferative disorders, including cancer, in
CC particular hepatocellular carcinoma, osteoclastoma, breast cancer,
CC or colon cancer. The products can also be used for detection and
CC diagnosis.
XX
SQ Sequence 175 AA;

Query Match 84.6%; Score 899; DB 19; Length 175;
Best Local Similarity 100.0%; Pred. No. 1.8e-90;
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 32 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 91
Db 1 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 60

Qy 92 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 151
Db 61 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 120

Qy 152 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 206
Db 121 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 175

RESULT 6
AAW37872
ID AAW37872 standard; Protein; 175 AA.
XX
AC AAW37872;
XX
DT 10-AUG-1998 (first entry)
XX
DE Human protein comprising secretory signal amino acid sequence 9.
XX
KW Human protein; secretory signal; nutritional source; cytokine;
KW immunity;haematopoiesis; activin; inhibin; tumour; chemotactic;
KW chemokinetic; thrombolytic; anti-inflammatory; inhibition;
KW stomach cancer cell.
XX
OS Homo sapiens.
XX
XX
PN WO9811217-A2.
XX
PD 19-MAR-1998.
XX
PF 12-SEP-1997; 97WO-JP03239.
XX
PR 13-SEP-1996; 96JP-0243060.
XX
PA (PROT-) PROTEGENE INC.
PA (SAGA) SAGAMI CHEM RES CENTRE.
XX
PI Kato S, Kobayashi M, Sekine S, Yamaguchi T;
XX
DR WPI; 1998-207380/18.
DR N-PSDB; AAV29047, AAV29048.
XX
PT Human proteins with secretory signal sequences - used to treat
PT immune deficiencies, infections, tumours, and haematopoietic
PT disorders, etc.
XX
PS Claim 1; Pages 79; 131pp; English.
XX
CC This is the amino acid sequence of a novel human protein comprising
CC a secretory signal isolated from stomach cancer cells. Its proteins
CC can be used as nutritional sources or supplements. The proteins may
CC also have cytokine functions, immune modulating functions,
CC haematopoiesis regulating activity, activin/inhibin regulating
CC activity, chemotactic/chemokinetic activity, haemostatic and
CC thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, tumour inhibition activity.
XX
SQ Sequence 175 AA;

Query Match 84.6%; Score 899; DB 19; Length 175;
Best Local Similarity 100.0%; Pred. No. 1.8e-90;
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 32 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 91
Db 1 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 60

Qy 92 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 151
Db 61 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 120

Qy 152 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 206
Db 121 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 175

RESULT 7
AAY59675
ID AAY59675 standard; Protein; 175 AA.

```
XX
AC   AAY59675;
XX
DT   18-JAN-2000  (first entry)
XX
DE   Secreted protein 108-008-5-0-A6-FL.
XX
KW   Secreted protein; fingerprint identification technique;
KW   chromosome mapping; human; hereditary disease; diagnosis; cancer;
KW   hyperlipidaemia; cardiovascular; neurodegenerative disorder; therapy;
KW   autoimmune disease; rheumatic disease; embryogenic disorder; myopathy;
KW   renal injury; amino aciduria; hypoglycaemia; male rat infertility;
KW   hypertension.
XX
OS   Homo sapiens.
XX
PN   WO9940189-A2.
XX
PD   12-AUG-1999.
XX
PF   09-FEB-1999;   99WO-IB00282.
XX
PR   09-FEB-1998;   98US-0074121.
PR   13-APR-1998;   98US-0081563.
PR   10-AUG-1998;   98US-0096116.
PR   04-SEP-1998;   98US-0099273.
XX
PA   (GEST ) GENSET.
XX
PI   Bougueleret L,  Duclert A,  Dumas Milne Edwards J;
XX
DR   WPI; 1999-600966/51.
DR   N-PSDB; AAZ40803.
XX
PT   Extended cDNAs useful for expressing secreted proteins and to obtain
PT   specific antibodies -
XX
PS   Claim 10; Page 199; 244pp; English.
XX
CC   This sequence represents a human secreted protein of the invention.
CC   The extended cDNAs (or genomic DNAs obtainable from them) may be used to
CC   prepare PCR primers and probes. These are useful for forensic matching or
CC   positive identification by DNA sequencing. They may also be used in
CC   alternative fingerprint identification techniques. Antibodies against the
CC   proteins encoded by the extended cDNAs are useful in identification of
CC   tissue types or cell species, as well as identifying tissue specific
CC   soluble proteins. The sequences can be used for chromosome mapping and
CC   identification of genes associated with hereditary diseases or drug
CC   response. signal sequences from the cDNAs can be used in construction of
CC   secretion vectors. Other sequences derived from the extended cDNAs can be
CC   used to clone upstream genomic DNA sequences including promoters. This is
CC   in turn useful for identifying proteins that interact with promoter
CC   sequences. Some of the proteins may be useful in diagnosing and treating
CC   several disorders including, but not limited to: cancer, hyperlipidaemia,
CC   cardiovascular and neurodegenerative disorders, autoimmune diseases, and
CC   rheumatic diseases, embryogenic disorders, hypertension, renal injury,
CC   amino acidurias, hypoglycaemia, male rat infertility and myopathies.
XX
SQ   Sequence   175 AA;

  Query Match             84.6%;  Score 899;  DB 20;  Length 175;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-90;
  Matches  175;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      32 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 91
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 60

Qy      92 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 151
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 120

Qy     152 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 206
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 175


RESULT 8
AAB24070
ID   AAB24070 standard; Protein; 175 AA.
XX
AC   AAB24070;
XX
DT   29-JAN-2001  (first entry)
XX
DE   Human PRO1030 protein sequence SEQ ID NO:40.
XX
KW   Human; tumour; diagnosis; neoplastic disease; neoplastic cell growth;
KW   proliferation; tumourigenesis; identification; cancer; cytostatic;
KW   nootropic; neuroprotective; antiinflammatory; immunosuppressive;
KW   immunostimulant; antiangiogenic; leukaemia; lymphoid malignancy;
KW   neuronal disorder; glial disorder; astrocytal disorder; angiogenic;
KW   hypothalamic disorder; glandular disorder; macrophagal disorder;
KW   epithelial disorder; stromal disorder; blastocoelic disorder;
KW   inflammatory disorder; immunologic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200053755-A2.
XX
PD   14-SEP-2000.
XX
PF   06-JAN-2000; 2000WO-US00376.
XX
PR   08-MAR-1999;   99WO-US05028.
PR   02-JUN-1999;   99WO-US12252.
PR   23-JUN-1999;   99US-0141037.
PR   07-JUL-1999;   99US-0143048.
PR   26-JUL-1999;   99US-0145698.
PR   30-NOV-1999;   99WO-US28313.
PR   20-DEC-1999;   99WO-US30911.
PR   05-JAN-2000; 2000WO-US00219.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Ashkenazi AJ,  Baker KP,  Goddard A,  Gurney AL,  Hillan KJ,  Roy MA;
PI   Watanabe CK,  Wood WI;
XX
DR   WPI; 2000-572270/53.
DR   N-PSDB; AAC58380.
XX
PT   Thirty PRO polynucleotides encoding PRO polypeptides, useful in the
PT   treatment, diagnosis and prevention of cancer -
XX
PS   Claim 61; Fig 28; 286pp; English.
XX
CC   The present invention describes an isolated antibody  that binds to
CC   one of the human PRO proteins designated PRO212, PRO290, PRO341, PRO535,
CC   PRO619, PRO717, PRO809, PRO830, PRO848, PRO943, PRO1005, PRO1009,
CC   PRO1025, PRO1030, PRO1097, PRO1107, PRO1111, PRO1153, PRO1182, PRO1184,
CC   PRO1187, PRO1281, PRO23, PRO39, PRO834, PRO1317, PRO1710, PRO2094,
CC   PRO2145 OR PRO2198. PRO antagonists can be used to inhibit tumour cell
CC   growth. The PRO polypeptides and nucleotides are useful in the
CC   treatment, diagnosis and prevention of cancer. The antibodies and other
CC   anti-tumour compounds maybe used to treat various conditions, including
CC   those characterised by overexpression and/or activation of the amplified
CC   PRO genes. Exemplary conditions or disorders to be treated with such
CC   antibodies and other compounds include benign or malignant tumours
CC   (e.g., renal, liver, kidney, bladder, breast, gastric, ovarian,
CC   colorectal, prostate, pancreatic, lung, vulva, thyroid, hepatic
CC   carcinomas, sarcomas, glioblastomas, and various head and neck tumours),
CC   leukaemias and lymphoid malignancies, other disorders such as neuronal,
CC   glial, astrocytal, hypothalamic and other glandular, macrophagal,
CC   epithelial, stromal and blastocoelic disorders, and inflammatory,
CC   angiogenic and immunologic disorders. AAC58242 to AAC58366 represent PCR
CC   primers and hybridisation probes used in the isolation of the human PRO
```

```
CC   sequences. AAC58367 to AAC58396 and AAB24057 to AAB24089 represent human
CC   PRO polynucleotide and protein sequences given in the exemplification of
CC   the present invention.
XX
SQ   Sequence   175 AA;

  Query Match             84.6%;  Score 899;  DB 21;  Length 175;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-90;
  Matches  175;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      32 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 91
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 60

Qy      92 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 151
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 120

Qy     152 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 206
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 175


RESULT 9
AAU08804
ID   AAU08804 standard; Protein; 175 AA.
XX
AC   AAU08804;
XX
DT   19-DEC-2001  (first entry)
XX
DE   Breast cancer cell membrane associated protein 7 (BCMP 7).
XX
KW   Breast cancer cell membrane protein 7; BCMP 7; breast cancer;
KW   metastasis; cytostatic; gene therapy; antibody; antisense;
KW   chromosome 7p21.3; vaccine.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..20
FT                   /note= "Signal peptide"
FT   Protein         21..175
FT                   /note= "Mature BCMP 7"
XX
PN   WO200163290-A1.
XX
PD   30-AUG-2001.
XX
PF   21-FEB-2001; 2001WO-GB00734.
XX
PR   25-FEB-2000; 2000GB-0004576.
XX
PA   (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI   Boyd RS,  Stamps AC,  Terrett JA,  Tyson KL;
XX
DR   WPI; 2001-570651/64.
DR   N-PSDB; AAS13480.
XX
PT   Diagnosing, preventing and treating breast cancer using a breast cancer
PT   cell membrane protein BCMP 7 -
XX
PS   Claim 1; Fig 1; 62pp; English.
XX
CC   The invention describes the novel use of a protein found in breast
CC   cancer cell membranes (BCMP 7) for diagnosing, preventing and treating
CC   breast cancers. The peptide has cytostatic action and potential uses in
CC   gene therapy and in vaccines. The polypeptide, antisense nucleic acids,
CC   or fusion proteins comprising and Green Fluorescent protein or the DsRed
CC   Fluorescent protein, antibodies specific for and/or nucleic acid are
CC   used for the prevention and/or treatment of breast cancer. Antibodies
CC   against may also be used for screening for and/or diagnosis of breast
CC   cancer in a patient. The method for monitoring/assessing breast cancer
CC   treatment in a patient and for the identification of metastatic breast
CC   cancer cells in samples from a patient. This sequence is breast
CC   cancer cell associated protein 7 (BCMP 7), encoded by a gene located on
CC   chromosome 7p21.3, described in the method of the invention.
XX
SQ   Sequence   175 AA;

  Query Match             84.6%;  Score 899;  DB 22;  Length 175;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-90;
  Matches  175;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      32 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 91
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 60

Qy      92 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 151
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 120

Qy     152 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 206
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 175


RESULT 10
AAB72203
ID   AAB72203 standard; Protein; 175 AA.
XX
AC   AAB72203;
XX
DT   04-MAY-2001  (first entry)
XX
DE   Human huXAG-1/CCSG colon cancer specific gene amino acid sequence.
XX
KW   Human; growth factor; huXAG-1; colon cancer specific gene; CCSG;
KW   cell proliferation; liver disease; fulminant liver failure; cirrhosis;
KW   hepatitis; cancer; colon cancer; colorectal carcinoma.
XX
OS   Homo sapiens.
XX
PN   US6171816-B1.
XX
PD   09-JAN-2001.
XX
PF   22-AUG-1997;   97US-0916576.
XX
PR   23-AUG-1996;   96US-0024347.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Yu G,  Dillon PJ,  Ebner R,  Endress GA;
XX
DR   WPI; 2001-136872/14.
DR   N-PSDB; AAF63314.
XX
PT   Novel human growth factor polypeptide useful for diagnosing and
PT   treating colon cancer and liver diseases, to prevent and heal damage to
PT   the lungs and for identifying modulators of therapeutic use -
XX
PS   Claim 165; Fig 1; 59pp; English.
XX
CC   This invention relates to a human growth factor polypeptide huXAG-1 also
CC   known as a colon cancer specific gene (CCSG). HuXAG-1 stimulates cell
CC   proliferation as a growth factor. The HuXAG-1 protein is useful for
CC   identifying compounds capable of enhancing or inhibiting cellular
CC   response induced by huXAG-1. The protein is also useful for stimulating
CC   proliferation of cells e.g. colon, breast, liver and lung cells, and
CC   hepatocytes. It is useful for alleviating or treating liver diseases and
CC   pathologies such as fulminant liver failure caused by cirrhosis, liver
CC   damage caused by viral hepatitis and toxic substances, for preventing and
CC   treating damage to the lungs caused by various pathological states and
CC   for promoting healing of breast tissue injury due to surgery, trauma or
```

```
CC   cancer. HuXAG-1 and the identified antagonist are useful for treating
CC   cancer, in particular colon cancer. Detecting altered levels of huXAG-1
CC   and its polynucleotides are useful for diagnosing or detecting cancer in
CC   mammals. The gene encoding huXAG-1 is useful for monitoring human
CC   colorectal carcinoma. huXAG-1 nucleic acid molecules are also useful for
CC   chromosome identification. The present sequence represents the huXAG-1
CC   protein.
XX
SQ   Sequence   175 AA;

  Query Match             84.6%;  Score 899;  DB 22;  Length 175;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-90;
  Matches  175;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      32 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 91
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 60

Qy      92 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 151
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 120

Qy     152 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 206
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 175


RESULT 11
AAG00109
ID   AAG00109 standard; Protein; 132 AA.
XX
AC   AAG00109;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein, SEQ ID NO: 4190.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   N-PSDB; AAC00115.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 13; SEQ ID 4190; 71pp + CD-ROM; English.
XX
CC   The present sequence is a polypeptide encoded by one of a large number
CC   of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC   were prepared from total human RNAs or polyA+ RNAs derived from 30
CC   different tissues. EST sequences usually correspond mainly to the 3'
CC   untranslated region (UTR) of the mRNA because they are often obtained
CC   from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC   isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC   those cases where longer cDNA sequences have been obtained, the full 5'
CC   UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC   ends and can therefore be used to obtain full length cDNAs and genomic
CC   DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC   chromosome mapping procedures. They are used to obtain upstream
CC   regulatory sequences and to design expression and secretion vectors.
XX
SQ   Sequence   132 AA;

  Query Match             61.7%;  Score 656;  DB 21;  Length 132;
  Best Local Similarity   100.0%;  Pred. No. 5.9e-64;
  Matches  126;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      32 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 91
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 60

Qy      92 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 151
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 120

Qy     152 DGQYVP 157
           ||||||
Db     121 DGQYVP 126


RESULT 12
AAY59718
ID   AAY59718 standard; Protein; 131 AA.
XX
AC   AAY59718;
XX
DT   18-JAN-2000  (first entry)
XX
DE   Secreted protein 78-21-1-B7-FL1.
XX
KW   Secreted protein; fingerprint identification technique;
KW   chromosome mapping; human; hereditary disease; diagnosis; cancer;
KW   hyperlipidaemia; cardiovascular; neurodegenerative disorder; therapy;
KW   autoimmune disease; rheumatic disease; embryogenic disorder; myopathy;
KW   renal injury; amino aciduria; hypoglycaemia; male rat infertility;
KW   hypertension.
XX
OS   Homo sapiens.
XX
PN   WO9940189-A2.
XX
PD   12-AUG-1999.
XX
PF   09-FEB-1999;   99WO-IB00282.
XX
PR   09-FEB-1998;   98US-0074121.
PR   13-APR-1998;   98US-0081563.
PR   10-AUG-1998;   98US-0096116.
PR   04-SEP-1998;   98US-0099273.
XX
PA   (GEST ) GENSET.
XX
PI   Bougueleret L,  Duclert A,  Dumas Milne Edwards J;
XX
DR   WPI; 1999-600966/51.
DR   N-PSDB; AAZ40846.
XX
PT   Extended cDNAs useful for expressing secreted proteins and to obtain
PT   specific antibodies -
XX
PS   Claim 10; Page 240; 244pp; English.
XX
CC   This sequence represents a human secreted protein of the invention.
CC   The extended cDNAs (or genomic DNAs obtainable from them) may be used to
CC   prepare PCR primers and probes. These are useful for forensic matching or
CC   positive identification by DNA sequencing. They may also be used in
CC   alternative fingerprint identification techniques. Antibodies against the
CC   proteins encoded by the extended cDNAs are useful in identification of
CC   tissue types or cell species, as well as identifying tissue specific
CC   soluble proteins. The sequences can be used for chromosome mapping and
```

```
CC   identification of genes associated with hereditary diseases or drug
CC   response. signal sequences from the cDNAs can be used in construction of
CC   secretion vectors. Other sequences derived from the extended cDNAs can be
CC   used to clone upstream genomic DNA sequences including promoters. This is
CC   in turn useful for identifying proteins that interact with promoter
CC   sequences. Some of the proteins may be useful in diagnosing and treating
CC   several disorders including, but not limited to: cancer, hyperlipidaemia,
CC   cardiovascular and neurodegenerative disorders, autoimmune diseases, and
CC   rheumatic diseases, embryogenic disorders, hypertension, renal injury,
CC   amino acidurias, hypoglycaemia, male rat infertility and myopathies.
XX
SQ   Sequence   131 AA;

  Query Match             56.9%;  Score 605;  DB 20;  Length 131;
  Best Local Similarity   92.9%;  Pred. No. 2.3e-58;
  Matches  117;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      32 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 91
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEX 60

Qy      92 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 151
             | ||||||||||||||| |||||||| ||||| ||||| ||| ||||||||||||||
Db      61 XLXKSKTSNKPLMIIHHLDXCPHSQALKKXFAENKXIQKLAXQFVXLNLVYETTDKHLSP 120

Qy     152 DGQYVP 157
           |||| |
Db     121 DGQYXP 126


RESULT 13
AAY12312
ID   AAY12312 standard; Protein; 116 AA.
XX
AC   AAY12312;
XX
DT   17-JUN-1999  (first entry)
XX
DE   Human 5' EST secreted protein SEQ ID NO:343.
XX
KW   Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW   forensic; gene therapy; chromosome mapping; signal peptide;
KW   upstream regulatory sequence; cytokine activity; cell proliferation;
KW   differentiation; haematopoiesis regulation; tissue growth regulation;
KW   reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; anti-inflammatory; tumour inhibition.
XX
OS   Homo sapiens.
XX
PN   WO9906548-A2.
XX
PD   11-FEB-1999.
XX
PF   31-JUL-1998;   98WO-IB01222.
XX
PR   01-AUG-1997;   97US-0905135.
XX
PA   (GEST ) GENSET.
XX
PI   Duclert A,  Dumas Milne Edwards J,  Lacroix B;
XX
DR   WPI; 1999-153778/13.
DR   N-PSDB; AAX41145.
XX
PT   New nucleic acids encoding human secreted proteins - obtained from
PT   cDNA libraries prepared from e.g. liver, ovary, brain, prostate,
PT   kidney, lung, umbilical cord, placenta and colon tissue
XX
PS   Claim 27; Page 682; 824pp; English.
XX
CC   AAX41094 to AAX41347 represent 5' expressed sequence tags (ESTs) for
CC   human secreted proteins, and encode the proteins given in AAY12261 to
CC   AAY12514, respectively. The proteins given represent the signal peptide
CC   and an N-terminal fragment of a secreted protein. The nucleic acid
CC   sequences can be used for producing secreted human gene products. They
CC   can also be used to develop products for diagnosis and therapy. The
CC   proteins obtained may have cytokine activity, cell
CC   proliferation/differentiation activity, haematopoiesis regulating
CC   activity, tissue growth regulating activity, reproductive hormone
CC   regulating activity, chemotactic/ chemokinetic activity, haemostatic and
CC   thrombolytic activity, receptor/ ligand activity, anti-inflammatory
CC   activity, tumour inhibition activity or other activities. The products
CC   can be used in forensic, gene therapy and chromosome mapping procedures.
CC   The sequences can also be used for obtaining corresponding promoter
CC   sequences. The nucleic acids encoding the signal peptide can be used for
CC   directing extracellular secretion of a polypeptide or the insertion of a
CC   polypeptide into a membrane, or importing a polypeptide into a cell.
XX
SQ   Sequence   116 AA;

  Query Match             56.3%;  Score 598;  DB 20;  Length 116;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-57;
  Matches  116;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      32 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 91
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 60

Qy      92 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDK 147
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDK 116


RESULT 14
AAY64672
ID   AAY64672 standard; Protein; 115 AA.
XX
AC   AAY64672;
XX
DT   01-FEB-2000  (first entry)
XX
DE   Human 5' EST related polypeptide SEQ ID NO:833.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
KW   gene therapy; chromosome mapping; upstream regulatory sequence;
KW   forensic; location; development; protein synthesis; stability;
KW   regulation; identification.
XX
OS   Homo sapiens.
XX
PN   WO9953051-A2.
XX
PD   21-OCT-1999.
XX
PF   09-APR-1999;   99WO-IB00712.
XX
PR   09-APR-1998;   98US-0057719.
PR   28-APR-1998;   98US-0069047.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-038446/03.
DR   N-PSDB; AAZ42286.
XX
PT   Novel secreted protein 5' expressed sequence tag sequences used in
PT   diagnostic, forensic, gene therapy, and chromosome mapping procedures
XX
PS   Claim 3; Page 604; 837pp; English.
XX
CC   AAZ42265 to AAZ43075 represent novel 5' expressed sequence tag (EST)
CC   sequences, corresponding to human secreted proteins. AAY64651 to
CC   AAY65438 represent the EST-related proteins corresponding to AAZ42265 to
CC   AAZ43052. The 5' ESTs can be used for producing secreted human gene
CC   products. They can be used to identify and isolate 5' untranslated
```

```
CC   regions (UTRs) and upstream regulatory regions which control the
CC   location, development stage, rate, and quantity of protein synthesis, as
CC   well as stability of mRNA. The ESTs are also useful as probes for
CC   chromosome mapping, and to obtain full length cDNA clones. The ESTs can
CC   also be used in forensic procedures to identify individuals, or in
CC   diagnostic procedures to identify individuals having genetic diseases
CC   resulting from abnormal gene expression. The products may also be used in
CC   gene therapy protocols. The nucleic acids encoding signal peptides can be
CC   used for directing extracellular secretion of a polypeptide or the
CC   insertion of a polypeptide into a membrane, or importing a polypeptide
CC   into a cell. The proteins encoded by the EST sequences may be useful in
CC   treating a variety of human conditions. Secreted proteins have
CC   therapeutic value, and the identification of new secreted proteins is
CC   valuable. AAZ42249 to AAZ42264 and AAY64644 to AAY64650 represent
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence   115 AA;

  Query Match             55.8%;  Score 593;  DB 21;  Length 115;
  Best Local Similarity   100.0%;  Pred. No. 4e-57;
  Matches  115;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      32 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 91
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 60

Qy      92 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTD 146
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTD 115


RESULT 15
AAW37846
ID   AAW37846 standard; Protein; 166 AA.
XX
AC   AAW37846;
XX
DT   28-AUG-1998  (first entry)
XX
DE   Human XAG growth factor huXAG-3.
XX
KW   HuXAG-3; XAG; growth factor; breast disease; liver disease;
KW   lung disease; emphysema; wound healing; cancer; diagnosis; therapy;
KW   human.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..23
FT                   /label= Sig_peptide
FT   Protein         24..166
FT                   /label= Mat_protein
FT                   /note= "Claim 11"
FT   Peptide         25..44
FT                   /label= Epitope
FT                   /note= "Claim 12"
FT   Peptide         104..115
FT                   /label= Epitope
FT                   /note= "Claim 12"
XX
PN   WO9807749-A1.
XX
PD   26-FEB-1998.
XX
PF   22-AUG-1997;   97WO-US14139.
XX
PR   23-AUG-1996;   96WO-US13766.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Dillon PJ,  Ebner R,  Endress GA,  Yu G;
XX
DR   WPI; 1998-169093/15.
DR   N-PSDB; AAV19157.
XX
PT   New isolated human XAG growth factor(s) - used to develop products
PT   for treating e.g. liver, lung or breast diseases or
PT   hyperproliferative disorders, e.g. cancer.
XX
PS   Claim 1; Fig 3A-B; 141pp; English.
XX
CC   This polypeptide comprises huXAG-3, a member of a novel family of
CC   human growth factors also including huXAG-1 (see AAW37844) and
CC   huXAG-2 (see AAW37845).  These proteins share homology with the
CC   XAG protein of Xenopus laevis, which is involved in embryogenesis
CC   and is expressed in adult tissue.  huXAG-3 cDNA (see AAV19157) was
CC   isolated from a human small intestine cDNA library.  Vectors, host
CC   cells, antibodies, and screening methods for identifying agonists
CC   and antagonists of huXAG-3 are provided.  HuXAG polypeptides are
CC   growth factors and can be used to stimulate proliferation of cells.
CC   They can be used to stimulate the proliferation and differentiation
CC   of hepatocytes to alleviate or treat liver diseases and pathologies
CC   such as fulminant liver failure caused by cirrhosis, liver damage
CC   caused by viral hepatitis and toxic substances.  They can also be
CC   used to stimulate or promote liver regeneration, e.g. after
CC   surgery.  They can also be used to prevent and heal damage to the
CC   lungs caused by various pathological states.  They can be used to
CC   stimulate proliferation and differentiation and promote the repair
CC   of alveoli and bronchiolar epithelium to prevent, attenuate, or
CC   treat acute or chronic lung damage, e.g. emphysema, which results
CC   in the progressive loss of alveoli, and inhalation injuries, e.g.
CC   resulting from smoke inhalation and burns, that cause necrosis of
CC   the bronchiolar epithelium and alveoli.  They can also be used to
CC   stimulate the proliferation and differentiation of breast tissue
CC   and could therefore be used to promote healing of breast tissue
CC   injury due to surgery, trauma or cancer.  Antagonists can be used
CC   to treat hyperproliferative disorders, including cancer, in
CC   particular hepatocellular carcinoma, osteoclastoma, breast cancer,
CC   or colon cancer.  The products can also be used for detection and
CC   diagnosis.
XX
SQ   Sequence   166 AA;

  Query Match             52.2%;  Score 555;  DB 19;  Length 166;
  Best Local Similarity   64.7%;  Pred. No. 1e-52;
  Matches  108;  Conservative   28;  Mismatches   19;  Indels   12;  Gaps    4;

Qy      41 LLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEEALYKSKTSN 100
           |||| :|  ||    :|       |: ||  |||||||||| : | ||||| |: :: |
Db      11 LLLVTVSSNLA--IAIK-------KEKRP--PQTLSRGWGDDITWVQTYEEGLFYAQKSK 59

Qy     101 KPLMIIHHLDECPHSQALKKVFAENKEIQKLAE-QFVLLNLVYETTDKHLSPDGQYVPRI 159
           ||||:||||::| :|||||||||:|:|||::|: :|::|||::|||||:||||||||||
Db      60 KPLMVIHHLEDCQYSQALKKVFAQNEEIQEMAQNKFIMLNLMHETTDKNLSPDGQYVPRI 119

Qy     160 MFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 206
           |||||||||||||| |||||||| ||| |  ||::||||||:|:::||
Db     120 MFVDPSLTVRADIAGRYSNRLYTYEPRDLPLLIENMKKALRLIQSEL 166


Search completed: October  9, 2002, 16:40:04
Job time : 33 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 16:39:22 ; Search time 21 Seconds
(without alignments)
942.591 Million cell updates/sec

Title: US-09-674-266A-181
Perfect score: 1063
Sequence: 1 RLSCAGTLSGSGPHPSRRLT.....DTALLLDNMKKALKLLKTEL 206

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	899	84.6	175	2	JE0350	Anterior gradient-
2	199	18.7	186	2	T31643	hypothetical prote
3	119	11.2	150	2	D69100	probable protein d
4	87.5	8.2	458	2	S50919	hypothetical prote
5	87	8.2	348	2	G72016	probable disulfide
6	87	8.2	348	2	C86607	disulfide bond iso
7	85.5	8.0	1830	1	S19188	myosin-V - chicken
8	84.5	7.9	1151	2	AG1717	probable peptidogl
9	84	7.9	771	2	S51421	hypothetical prote
10	83.5	7.9	217	2	T37859	probable transcrip
11	83.5	7.9	918	2	C96829	unknown protein F1
12	83	7.8	479	2	T40683	cell cycle protein
13	83	7.8	507	2	S56143	cell cycle protein
14	82.5	7.8	788	2	C70441	lipoate-protein li
15	81.5	7.7	970	2	S29069	inositol polyphosp
16	81.5	7.7	1158	2	T50454	probable rho1 GDP-
17	81.5	7.7	1397	2	T10466	DNA topoisomerase
18	81	7.6	610	2	T22909	hypothetical prote
19	81	7.6	635	2	T23874	hypothetical prote
20	80	7.5	154	2	F82633	probable protein d
21	80	7.5	259	2	A96740	F14O23.14 [importe
22	80	7.5	843	2	S53947	hypothetical prote
23	79.5	7.5	277	2	T37629	hypothetical prote
24	79.5	7.5	301	2	G71929	GTP-binding protei
25	79.5	7.5	642	2	F83718	ABC transporter (A
26	79.5	7.5	921	2	F84593	hypothetical prote
27	79	7.4	477	2	T37791	probable rna polym
28	79	7.4	510	1	S43516	carboxypeptidase C
29	78.5	7.4	280	2	A69590	shikimate 5-dehydr
30	78.5	7.4	917	2	T05430	hypothetical prote
31	78.5	7.4	1033	2	A96714	hypothetical prote
32	78.5	7.4	1262	2	I48855	tripeptidyl-peptid
33	78.5	7.4	1570	2	AC2012	hypothetical prote
34	78	7.3	337	2	S06956	segmentation prote
35	78	7.3	1828	2	B59254	mysoin heavy chain
36	78	7.3	1853	1	A46761	myosin heavy chain
37	78	7.3	1855	2	A59254	mysoin heavy chain
38	76.5	7.2	985	2	E69850	formate dehydrogen
39	76.5	7.2	1025	2	T10259	RNA-directed DNA p
40	76	7.1	166	2	E72018	thioredoxin disulf
41	76	7.1	166	2	D86606	thioredoxin disulf
42	76	7.1	367	2	T39574	probable uracil ki
43	76	7.1	543	1	F64871	oligopeptide-bindi
44	76	7.1	583	2	E75529	probable peptide A
45	76	7.1	1651	2	F88750	protein vit-6 [imp

ALIGNMENTS

RESULT 1
JE0350
Anterior gradient-2 - human
C;Species: Homo sapiens (man)
C;Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 21-Jul-2000
C;Accession: JE0350
R;Thompson, D.A.; Weigel, R.J.
Biochem. Biophys. Res. Commun. 251, 111-116, 1998
A;Title: hAG-2, the human homologue of the Xenopus laevis cement gland gene XAG-2, is
A;Reference number: JE0350; MUID:99009231
A;Accession: JE0350
A;Molecule type: mRNA
A;Residues: 1-175 <THO>
A;Cross-references: GB:AF007791; NID:g3779196; PIDN:AAC77358.1; PID:g3779197
C;Comment: This protein is coexpressed with estrogen receptor (ER) in breast cancer c
C;Genetics:
A;Gene: hAG-2
C;Superfamily: human anterior gradient-2

Query Match 84.6%; Score 899; DB 2; Length 175;
Best Local Similarity 100.0%; Pred. No. 3.2e-74;
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      32 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 91
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 60

Qy      92 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 151
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 120

Qy     152 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 206
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 175
```

RESULT 2
T31643
hypothetical protein Y57A10A.u - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 20-Jun-2000
C;Accession: T31643
R;Smye, R.
submitted to the EMBL Data Library, September 1999
A;Reference number: Z21048
A;Accession: T31643
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-186 <WIL>
A;Cross-references: EMBL:AL117195; PIDN:CAB55026.1; CESP:Y57A10A.u
A;Experimental source: clone Y57A10A
C;Genetics:

```
A;Gene: CESP:Y57A10A.u
A;Introns: 70/1
C;Superfamily: human anterior gradient-2

  Query Match             18.7%;  Score 199;  DB 2;  Length 186;
  Best Local Similarity   30.0%;  Pred. No. 1.1e-10;
  Matches   51;  Conservative   38;  Mismatches   57;  Indels   24;  Gaps    6;

Qy      41 LLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEEALYKSKTSN 100
           |||:||         |   |       :   :   |:||:|| : | : :|:|:  :  ::
Db       4 LLLLAL---------VSASAYASFDKIKDSIQNPLARGFGDDIAWVK-WEDAIETALDTD 53

Qy     101 KPLMIIHHLDECPHSQALKKVFAEN---KEIQKLAEQFVLLNLV--YETTDKHLSPDGQY 155
           ||: :: |   |   :|||| | ::   |  :||:| ||::|     |  ::   |||:|
Db      54 KPIFLLIHKSWCHACKALKKTFQQSNAKKAFKKLSEHFVMVNTEDDDEPFEEEYRPDGKY 113

Qy     156 VPRIMFVDPSLTVRADITGRYSNRL-----YAYEPADTALLLDNMKKALK 200
           :|||::|:| :     |:   : |:      |||  :  | :|::||  ||
Db     114 IPRLLFLDKN----GDLLQEFKNKKAEYKNYAYYYSSPADILNSMKDVLK 159


RESULT 3
D69100
probable protein disulfide-isomerase (EC 5.3.4.1) MTH1745 [similarity] - Methanobacteriu
C;Species: Methanobacterium thermoautotrophicum
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Sep-2000
C;Accession: D69100
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
A;Reference number: A69000; MUID:98037514
A;Accession: D69100
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-150 <MTH>
A;Cross-references: GB:AE000930; GB:AE000666; NID:g2622872; PIDN:AAB86215.1; PID:g262287
A;Experimental source: strain Delta H
C;Genetics:
A;Gene: MTH1745
C;Keywords: intramolecular oxidoreductase; isomerase

  Query Match             11.2%;  Score 119;  DB 2;  Length 150;
  Best Local Similarity   27.1%;  Pred. No. 0.0016;
  Matches   35;  Conservative   28;  Mismatches   46;  Indels   20;  Gaps    4;

Qy      39 AFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEEALYKSKT 98
           | | || : ||:                  :|::||:|:    | | |   ::||: ::
Db      19 AGLSLVLIIYTV----------------QPRVPQSLTTDEKD-LKWYTEHDEAIKEASR 60

Qy      99 SNKPLMIIHHLDECPHSQALKKVFAENKEIQ-KLAEQFVLLNLVYETTDKHLSPDGQY-V 156
           : | : ::     ||  | |:    :| |:| :||| |: : :  :|:    |    | |
Db      61 TGKNVFMVFSASWCPACQKLESETLQNTEVQRRLAEDFIAVKIDVDTSPALSSRYRIYGV 120

Qy     157 PRIMFVDPS 165
           | :: :|||
Db     121 PTVIILDPS 129


RESULT 4
S50919
hypothetical protein YMR191w - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein YM9646.03
C;Species: Saccharomyces cerevisiae
C;Date: 10-Feb-1995 #sequence_revision 12-May-1995 #text_change 04-Mar-2000
C;Accession: S50919; S55721
R;Pearson, D.; Bowman, S.
submitted to the EMBL Data Library, January 1995
A;Reference number: S50917
A;Accession: S50919
A;Molecule type: DNA
A;Residues: 1-458 <PEA>
A;Cross-references: EMBL:Z47815; NID:g642280; PIDN:CAA87812.1; PID:g642283; MIPS:YMR1
R;Gangloff, S.; McDonald, J.P.; Bendixen, C.; Arthur, L.; Rothstein, R.
Mol. Cell. Biol. 14, 8391-8398, 1994
A;Title: The yeast type I topoisomerase Top3 interacts with Sgs1, a DNA helicase homo
A;Reference number: A56359; MUID:95059068
A;Accession: S55721
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-127,'Y',129-131,'LALG',136-249 <GAN>
A;Cross-references: EMBL:U22341; NID:g726277; PIDN:AAB60288.1; PID:g726278
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1995
C;Genetics:
A;Map position: 13R
C;Superfamily: Saccharomyces cerevisiae hypothetical protein YMR191w

  Query Match             8.2%;  Score 87.5;  DB 2;  Length 458;
  Best Local Similarity   20.1%;  Pred. No. 4.5;
  Matches   55;  Conservative   35;  Mismatches   74;  Indels  109;  Gaps   13;

Qy      15 PSRRLTQG-----RWVRKSRVAMEKIPVSAFLLLVALS---------------------- 47
           |:| :  |       |:| ||| : :|  |    |::||
Db      82 PTRNMAVGGNNWSMWLRMSRVHLRQITKSLDRTLISLSHGNFSHQYNRNIFVTWWKSLFE 141

Qy      48 ----YTLARDTTVKPGAKKDTK-------------DSRPKLPQTLSRG----W----GD 81
                :  |   || |  ::                  | |::|:   ||    |
Db     142 ASTAFRRASGLTVSPLTRRGIARFDHFRPVPNVSKFASFPRVPKGAPRGLFTNWNMTTSK 201

Qy      82 QLIWTQTYEEALYK---SKTSNKPLMI------IHHLDECPHSQ---------------- 116
           :|:  : |  :  |       :|  : :       :  |::| ||
Db     202 RLLGQRAYSTSSIKFTQEAVNNMTISLRCFFNSLGGLNQCSHSNSCKAYQNASNVTSKQD 261

Qy     117 -----ALKKV------FAENKEIQKLAEQFVLLNLVYETTDKHLSPDGQYV--------- 156
                ||||:       |  | |: |: :      :| ||:  ::   | |:
Db     262 HVQPVALKKLSQKDINFIRNLELFKIMK--TQNEVVDETSAYYMEKPGSYIEFTISEFNV 319

Qy     157 -----PRIMFVDPSLTVRADITGRYSNRLYAYE 184
                  : |:||||   ||:     |  | ||
Db     320 NGTFSAPLSFLDPSLL--ADLDEMIRN--YKYE 348


RESULT 5
G72016
probable disulfide bond isomerase - Chlamydophila pneumoniae (strains CWL029 and AR39
C;Species: Chlamydophila pneumoniae, Chlamydia pneumoniae
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
C;Accession: G72016; B81523
R;Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood,
Nature Genet. 21, 385-389, 1999
A;Title: Comparative genomes of Clamydia pneumoniae and C. trachomatis.
A;Reference number: A72000; MUID:99206606
A;Accession: G72016
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-348 <ARN>
A;Cross-references: GB:AE001673; GB:AE001363; NID:g4377252; PIDN:AAD19071.1; PID:g437
A;Experimental source: strain CWL029
R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe
Nucleic Acids Res. 28, 1397-1406, 2000
A;Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39
A;Reference number: A81500; MUID:20150255
A;Accession: B81523
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-348 <REA>
A;Cross-references: GB:AE002251; GB:AE002161; NID:g7189833; PIDN:AAF38711.1; PID:g718
A;Experimental source: strain AR39, HL cells
C;Genetics:
A;Gene: CPn0933; CP0928

  Query Match             8.2%;  Score 87;  DB 2;  Length 348;
  Best Local Similarity   25.0%;  Pred. No. 3.5;
  Matches   29;  Conservative   26;  Mismatches   37;  Indels   24;  Gaps    4;
```

```
Qy      29 RVAMEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQT 88
           | ::::: | | ||   || :|    |::    :|:                | ::|
Db       9 RCSLKQLKVLATLL---LSLSL---PTLEAAENRDS---------------DSIVWHLD 46

Qy      89 YEEALYKSKTSNKPLMIIHHLDEC--PHSQALKKVFAENKEIQKLAEQFVLLNLVY 142
           |:||| ||| :  ||::|    :   |  :  |:|    : |:::  :|| : : |
Db      47 YQEALQKSKEAELPLLVIFSGSDWNGPCMKIRKEVLESPEFIKRVQGKFVCVEVEY 102


RESULT 6
C86607
disulfide bond isomerase [imported] - Chlamydophila pneumoniae (strain J138)
C;Species: Chlamydophila pneumoniae, Chlamydia pneumoniae
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
C;Accession: C86607
R;Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Is
Nucleic Acids Res. 28, 2311-2314, 2000
A;Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
A;Reference number: A86491; MUID:20330349
A;Accession: C86607
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-348 <STO>
A;Cross-references: GB:BA000008; NID:g8979307; PIDN:BAA99141.1; GSPDB:GN00142
A;Experimental source: strain J138
C;Genetics:
A;Gene: CPj0933

  Query Match             8.2%;  Score 87;  DB 2;  Length 348;
  Best Local Similarity   25.0%;  Pred. No. 3.5;
  Matches   29;  Conservative   26;  Mismatches   37;  Indels   24;  Gaps    4;

Qy      29 RVAMEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQT 88
           | ::::: | | ||   || :|    |::    :|:                | ::|
Db       9 RCSLKQLKVLATLL---LSLSL---PTLEAAENRDS---------------DSIVWHLD 46

Qy      89 YEEALYKSKTSNKPLMIIHHLDEC--PHSQALKKVFAENKEIQKLAEQFVLLNLVY 142
           |:||| ||| :  ||::|    :   |  :  |:|    : |:::  :|| : : |
Db      47 YQEALQKSKEAELPLLVIFSGSDWNGPCMKIRKEVLESPEFIKRVQGKFVCVEVEY 102


RESULT 7
S19188
myosin-V - chicken
N;Alternate names: calmodulin-binding protein; myosin I heavy chain, 190K
C;Species: Gallus gallus (chicken)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 19-Jan-2001
C;Accession: S19188; A44359; B44359; S29249
R;Espreafico, E.M.; Cheney, R.E.; Matteoli, M.; Nascimento, A.A.; De-Camilli, P.V.; Lars
submitted to the EMBL Data Library, February 1992
A;Description: Complete cDNA coding sequence of chicken brain p190, a calmodulin binding
A;Reference number: S19188
A;Accession: S19188
A;Molecule type: mRNA
A;Residues: 1-1830 <ESP>
A;Cross-references: EMBL:Z11718
R;Espreafico, E.M.; Cheney, R.E.; Matteoli, M.; Nascimento, A.A.; De Camilli, P.V.; Lars
J. Cell Biol. 119, 1541-1557, 1992
A;Title: Primary structure and cellular localization of chicken brain myosin-V (p190), a
A;Reference number: A44359; MUID:93107155
A;Accession: A44359
A;Molecule type: mRNA
A;Residues: 1-1688,'R',1690-1830 <ES2>
A;Experimental source: brain
A;Note: sequence extracted from NCBI backbone (NCBIN:121153, NCBIP:121154)
A;Note: the codon CGC for residue 1689 is inconsistent with the codon AGC for residue 16
A;Accession: B44359
A;Molecule type: protein
A;Residues: 155-164 <ES3>
R;Sanders, G.; Lichte, B.; Meyer, H.E.; Kilimann, M.W.
FEBS Lett. 311, 295-298, 1992
A;Title: cDNA encoding the chicken ortholog of the mouse dilute gene product. Sequence d
A;Reference number: S29249; MUID:93012002
A;Accession: S29249
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-1142,1144-1830 <SAN>
A;Cross-references: EMBL:X67251; NID:g63364; PIDN:CAA47673.1; PID:g63365
C;Comment: The neck domain comprises six approximately 23-residue tandem repeats; thi
C;Superfamily: myosin MYO2; myosin motor domain homology
C;Keywords: actin binding; ATP; brain; coiled coil; nucleotide binding; P-loop; phosp
F;72-752/Domain: myosin motor domain homology <MMOT>
F;163-170/Region: nucleotide-binding motif A (P-loop)
F;645-666/Region: actin binding #status predicted
F;765-909/Domain: neck #status predicted <NEC>
F;912-1420/Domain: coiled coil #status predicted <COI>
F;1421-1830/Domain: carboxyl-terminal domain #status predicted <CTD>
F;169/Binding site: ATP (Lys) #status predicted
F;1735/Binding site: phosphate (Thr) (covalent) #status predicted

  Query Match             8.0%;  Score 85.5;  DB 1;  Length 1830;
  Best Local Similarity   28.1%;  Pred. No. 38;
  Matches   27;  Conservative   21;  Mismatches   41;  Indels    7;  Gaps    3;

Qy      63 DTKDSRPKLPQTLSRGWGDQLIWTQTYEEALY-KSKTSNKPLMIIHHLDECPH--SQALK 119
           |  |   |:|:     |  :|  |   : ||: | : |||  :| |  |:  :     |:
Db     507 DLLDEECKMPKGSDDTWAQKLYNTHLNKCALFEKPRLSNKAFIIKHFADKVEYQCEGFLE 566

Qy     120 K----VFAENKEIQKLAEQFVLLNLVYETTDKHLSP 151
           |    |: |  :: | :::| ||  :::  :| :||
Db     567 KNKDTVYEEQIKVLKSSKKFKLLPELFQDEEKAISP 602


RESULT 8
AG1717
probable peptidoglycan bound protein (LPXTG motif) lin2283 [imported] - Listeria inno
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C;Accession: AG1717
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi,
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.;
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AG1717
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1151 <GLA>
A;Cross-references: GB:AL592022; PIDN:CAC97511.1; PID:g16414795; GSPDB:GN00178
A;Experimental source: strain Clip11262
C;Genetics:
A;Gene: lin2283

  Query Match             7.9%;  Score 84.5;  DB 2;  Length 1151;
  Best Local Similarity   22.8%;  Pred. No. 26;
  Matches   44;  Conservative   31;  Mismatches   51;  Indels   67;  Gaps   10;

Qy      50 LARDTTVKPGAKK--DTKDSRPKLPQTLSRGWGDQLIWTQ----------------TYE 90
           :| | :| |   :  |    :   |:||:  :|::   |                 |:
Db     961 VALDDSVAPVTVQYVDQNHKQIASPETLTGAYGEKFTAKQKKITNYTLVKTPANVSGTFN 1020

Qy      91 EA------LYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVL---LNLV 141
           |       :|:  |:    :|: ::|               |  :|||:  ||   ||
Db    1021 EKAQTVTFVYQKVTAGN--IIVDYVD--------------KNGEKLADSIVLTGKLNSS 1063

Qy     142 YETTDKHLSPDGQY-VPRIMFVDPSLTVRADITGRYSNR----LYAYEPADTALL----- 191
           | |: | :|    |  |:           : :|::||      | || | |  :
Db    1064 YRTSAKKISGYKLYQTPK-----------NASGKFSNTSQRVTYVYEKASTFSISSNKG 1111

Qy     192 LDNMKKALKLLKT 204
             ||::|: || :|
Db    1112 ADNVEKSAKLPQT 1124
```

RESULT 9
S51421
hypothetical protein YLR176c - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein L9470.18
C;Species: Saccharomyces cerevisiae
C;Date: 23-Feb-1995 #sequence_revision 12-May-1995 #text_change 06-Feb-1998
C;Accession: S51421
R;Wohldmann, P.
submitted to the EMBL Data Library, November 1994
A;Description: The sequence of S. cerevisiae cosmid 9470.
A;Reference number: S51414
A;Accession: S51421
A;Molecule type: DNA
A;Residues: 1-771 <WOH>
A;Cross-references: EMBL:U17246; NID:g577192; PID:g577210; MIPS:YLR176c
C;Genetics:
A;Gene: SGD:RFX1
A;Cross-references: SGD:S0004166; MIPS:YLR176c
A;Map position: 12R

Query Match 7.9%; Score 84; DB 2; Length 771;
Best Local Similarity 25.7%; Pred. No. 18;
Matches 47; Conservative 29; Mismatches 67; Indels 40; Gaps 8;

```
Qy       8 LSGSGP-----------HPSRRLTQGR----WVRKSRVAMEKIPVSAFLLLVALSYTLAR 52
           | ||||            |::  || :     |  : ::: :   ||: |      ||
Db      89 LIGSGPGSHKPSPTPTQPPAQPATQRQPATYSVYPASISLNRSNSSAYPLSFKSEETLNN 148

Qy      53 D--TTVK--------PGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEEALYKSKTSNKP 102
           :  |  |        | : |  | |: |   ::||    ::|  |  |    |||| :
Db     149 NPPTAAKRTNTFPSIPSSTKKQKTSQEKRISSISRRNTQEIIAKQIAEN--NKSKTIEEY 206

Qy     103 LMIIHH-------LDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSPDGQY 155
             |: |       :|    |:|  ::  :|:| ::  : | || |:     :| |    |
Db     207 AQIVKHAEIKVLSMDSQNTSKAALQLAEQNRERER--QVFALLWLMKNCKSQHDS----Y 260

Qy     156 VPR 158
           |||
Db     261 VPR 263
```

RESULT 10
T37859
probable transcription factor - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T37859
R;Devlin, K.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, February 1996
A;Reference number: Z21750
A;Accession: T37859
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-217 <DEV>
A;Cross-references: EMBL:Z69795; PIDN:CAA93690.1; GSPDB:GN00066; SPDB:SPAC17G8.07
A;Experimental source: strain 972h-; cosmid c17G8
C;Genetics:
A;Gene: SPDB:SPAC17G8.07
A;Map position: 1

Query Match 7.9%; Score 83.5; DB 2; Length 217;
Best Local Similarity 20.6%; Pred. No. 4.1;
Matches 42; Conservative 33; Mismatches 80; Indels 49; Gaps 8;

```
Qy      14 HPSRRLTQG-------RWVRKSRVAMEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKD 66
           |  |   :|       :||||          | |          ||   |     | :
Db      39 HTWRIFVEGVDGEDISKWVRK----------VVFKL---------HDTYNNP---TRTIE 76

Qy      67 SRPKLPQTLSRGWGDQLIWTQTYEEALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENK 126
           |  |   : :  |||:  |  : :    :  :   | |   |||   |:   :::: |
Db      77 SPPF--EVIETGWGEFDIMVRIF----FAPEAHEKALTFYHHLKLHPYGPRMEEMKASGG 130

Qy     127 EIQKLAEQFVLLNLVYETTDKHLSP----DGQYVPRIMFVDPSLTVRADITGRYSNRLYA 182
           :: :  : :: |  :| | | ||     ||            | | ::    :| :|
Db     131 LVESVQYEEIVFNEPFEYTYKLLSQNPIGDGH----------GLAVESEPDHPFSQQLEQ 180

Qy     183 YEPADTALLLDNMKKALKLLKTEL 206
           |      :  :|| ::: | ::
Db     181 DEADKLDFAIQEVKKTIEMYKQQV 204
```

RESULT 11
C96829
unknown protein F19K16.21 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: C96829
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719
A;Accession: C96829
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-918 <STO>
A;Cross-references: GB:AE005173; NID:g6453864; PIDN:AAF09048.1; GSPDB:GN00141
C;Genetics:
A;Gene: F19K16.21
A;Map position: 1

Query Match 7.9%; Score 83.5; DB 2; Length 918;
Best Local Similarity 21.3%; Pred. No. 24;
Matches 46; Conservative 41; Mismatches 76; Indels 53; Gaps 9;

```
Qy      36 PVSAFLLLVALSYTL------ARDTTVKPGAKKD---TKDSRPKLPQTLSRGWGDQLIWT 86
           | |  ||  :  |:|         |:::|  | :   ::||:|: |:  |     |   :
Db     149 PQSQILLEESSEYSLQTPESSGYKTSLQPNEKLEMTASQDSQPEQPK--SEAEESQPEDS 206

Qy      87 QTYEEALYKSKTSNKPLM----IIHHLDECPHSQALKKVFAENKEIQKLAEQFVLL---- 138
           :  |  :    | : |::     | ::||  : |   ::  || | :  :: | :
Db     207 EAKEVTVENKDTVHSPVLDGQHKITYMDETTNEQ---EILGENLEGRTSSKNFEVSPDIN 263

Qy     139 ------------NLVYETTDKHLSPDGQYVPR----------IMFVDPSLTVRADI--- 172
                       :|::|:     ||    :|:          : ||   :  | |
Db     264 HVNRIESPVAHPSLIFESDG---SPYESSIPKRSSSDEISERIVDFVSREIDSRLDTSEL 320

Qy     173 --TGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 206
             : | |:     : ||  | |:  || :|:|:  |
Db     321 NESQRSSSATNVSDSADVILELEKTKKEIKMLENAL 356
```

RESULT 12
T40683
cell cycle protein kinase hsk1 - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Jan-2000
C;Accession: T40683
R;Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Wedler, H.; Kutzner, M.; Wambutt, R.
submitted to the EMBL Data Library, January 1999
A;Reference number: Z21944
A;Accession: T40683
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-479 <LYN>
A;Cross-references: EMBL:AL035263; PIDN:CAB38417.1; GSPDB:GN00067; SPDB:SPBC776.12c
A;Experimental source: strain 972h-; cosmid c776
C;Genetics:

```
A;Gene: SPDB:SPBC776.12c
A;Map position: 2
A;Introns: 77/1; 92/2; 331/1
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo

  Query Match             7.8%;  Score 83;  DB 2;  Length 479;
  Best Local Similarity   22.4%;  Pred. No. 12;
  Matches   28;  Conservative   24;  Mismatches   55;  Indels   18;  Gaps    5;

Qy      83 LIWTQTYEEALYKSKTSNKP----LMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLL 138
           ::|      ::|| |  :||     : :  |::|      |:: ||    : |   |: |
Db     352 ILWASCGSASIYKEKLRHKPSQEERLCLDFLEKCLELDCNKRISAE----EALDHDFLYL 407

Qy     139 -NLVYETTDKHLSPDGQYVPRIMFVDPSLTVRADITGRYSNRLYAY---EPADTALLLDN 194
           || ||  |   : |       | : |     |:| :: : :  :   |  |    |
Db     408 DNLAYEKKDDDTAFDNS------FGETSFEKDEDLTAKHLSHILDFKEQEETDEPTSLSK 461

Qy     195 MKKAL 199
           |:::
Db     462 RKRSI 466


RESULT 13
S56143
cell cycle protein kinase hsk1 (EC 2.7.1.-) - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 28-Oct-1995 #sequence_revision 10-Nov-1995 #text_change 10-Dec-1999
C;Accession: S56143; T40682
R;Masai, H.; Miyake, T.; Arai, K.
EMBO J. 14, 3094-3104, 1995
A;Title: hsk1(+), a Schizosaccharomyces pombe gene related to Saccharomyces cerevisiae C
A;Reference number: S56143; MUID:95347336
A;Accession: S56143
A;Molecule type: DNA
A;Residues: 1-507 <MAS>
A;Cross-references: GB:D50493; NID:g807845; PIDN:BAA09087.1; PID:d1009722; PID:g807846
A;Experimental source: strain JY2
R;Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Wedler, H.; Kutzner, M.; Wambutt, R.
submitted to the EMBL Data Library, January 1999
A;Reference number: Z21944
A;Accession: T40682
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-507 <LYN>
A;Cross-references: EMBL:AL035263; PIDN:CAA22885.1; GSPDB:GN00067; SPDB:SPBC776.12c
A;Experimental source: strain 972h-; cosmid c776
C;Genetics:
A;Gene: hsk1; SPBC776.12c
A;Map position: 2
A;Introns: 77/1; 359/1
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo
C;Keywords: phosphotransferase; protein kinase
F;66-404/Domain: protein kinase homology <KIN>

  Query Match             7.8%;  Score 83;  DB 2;  Length 507;
  Best Local Similarity   22.4%;  Pred. No. 13;
  Matches   28;  Conservative   24;  Mismatches   55;  Indels   18;  Gaps    5;

Qy      83 LIWTQTYEEALYKSKTSNKP----LMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLL 138
           ::|      ::|| |  :||     : :  |::|      |:: ||    : |   |: |
Db     380 ILWASCGSASIYKEKLRHKPSQEERLCLDFLEKCLELDCNKRISAE----EALDHDFLYL 435

Qy     139 -NLVYETTDKHLSPDGQYVPRIMFVDPSLTVRADITGRYSNRLYAY---EPADTALLLDN 194
           || ||  |   : |       | : |     |:| :: : :  :   |  |    |
Db     436 DNLAYEKKDDDTAFDNS------FGETSFEKDEDLTAKHLSHILDFKEQEETDEPTSLSK 489

Qy     195 MKKAL 199
           |:::
Db     490 RKRSI 494


RESULT 14
C70441
lipoate-protein ligase A - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999
C;Accession: C70441
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;
V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666
A;Accession: C70441
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-788 <AQF>
A;Cross-references: GB:AE000748; NID:g2983960; PIDN:AAC07512.1; PID:g2983965; GB:AE00
A;Experimental source: strain VF5
C;Genetics:
A;Gene: lplA

  Query Match             7.8%;  Score 82.5;  DB 2;  Length 788;
  Best Local Similarity   23.0%;  Pred. No. 25;
  Matches   49;  Conservative   34;  Mismatches   73;  Indels   57;  Gaps   10;

Qy      25 VRKSRVAMEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSR-PKLPQTLSR-GW--G 80
           |:| : | : :|       :  :| : :   | || ::|   || | : :| :  |:  |
Db     585 VQKRQSAQKPVPNENLKEFIPQNYKVPK---VAPGPQEDM--SRLPIVRETSTEIGFING 639

Qy      81 DQLIWTQTYEEALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNL 140
           :: : :  : |:     : :| :|:  |           ::| |:| :|  |    |
Db     640 QKVSFEEALQVAVELLSKAKRPTLIVGPL----------VLWAWNEEAKKKGEIIKRLKE 689

Qy     141 VYETTDKHLSPDGQYVPRIMFVD---------PSLTV----------------RADITGR 175
           |   : || ||  | |:   ||          | |:|                | | | |
Db     690 KYPNLNIHLLPD--YKPKNKKVDFSREVDPPNPHLSVLYGEHDLTLMIGVHCYRTDFTIR 747

Qy     176 YSNR-----------LYAYEPADTALLLDNMKK 197
           :              || :  || :|   | ||
Db     748 MLKKYTDTKIVTLCTLYGHPDADVSLAGVNSKK 780


RESULT 15
S29069
inositol polyphosphate-5-phosphatase homolog - human
C;Species: Homo sapiens (man)
C;Date: 25-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 01-Dec-2000
C;Accession: S29069; I68621
R;Attree, O.; Olivos, I.M.; Okabe, I.; Bailey, L.C.; Nelson, D.L.; Lewis, R.A.; McInn
Nature 358, 239-242, 1992
A;Title: The Lowe's oculocerebrorenal syndrome gene encodes a protein highly homologo
A;Reference number: S29069; MUID:92334430
A;Accession: S29069
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-970 <ATT>
R;Leahey, A.M.; Charnas, L.R.; Nussbaum, R.L.
Hum. Mol. Genet. 2, 461-463, 1993
A;Title: Nonsense mutations in the OCRL-1 gene in patients with the oculocerebrorenal
A;Reference number: I54349; MUID:93278398
A;Accession: I68621
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 883-912 <RES>
A;Cross-references: GB:S62085; NID:g385336; PIDN:AAB26926.1; PID:g385337
C;Genetics:
A;Gene: OCRL-1

  Query Match             7.7%;  Score 81.5;  DB 2;  Length 970;
  Best Local Similarity   23.7%;  Pred. No. 40;
  Matches   45;  Conservative   32;  Mismatches   46;  Indels   67;  Gaps   13;

Qy      55 TVKPGAKKDTKDSRPK--LPQTLSRGWGDQLIWTQTYEEAL-YKS----KTS-NKPLMII 106
           | |  :| |  ||  |  :|      | |:::|  |    | |:|    ||| :||:  :
Db     553 TYKYDSKTDRWDSSGKCRVP-----AWCDRILWRGTNVNQLNYRSHMELKTSDHKPVSAL 607
```

```
Qy     107 HHL-----DECPHSQALKKVFAENKEIQKLAE-----------QFVLLNLVYETTDK-- 147
           |:     ||    :  :||| ::  |    |            :||  |: :    |
Db     608 FHIGVKVVDE----RRYRKVFEDSVRIMDRMENDFLPSLELSRREFVFENVKFRQLQKGK 663

Qy     148 -HLSPDGQ------YVPRIM---------------FVDPSLTVRADITGRYSNRLYAYE 184
             :| :||      ::|::                :::|: ||  ||:      |  |
Db     664 FQISNNGQVPCHFSFIPKLNDSQYCKPWLRAEPFEGYLEPNETV--DIS------LDVYV 715

Qy     185 PADTALLLDN 194
           |:  :|::
Db     716 SKDSVTILNS 725
```

Search completed: October 9, 2002, 16:42:16
Job time : 23 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 16:39:17 ; Search time 13 Seconds
(without alignments)
613.556 Million cell updates/sec

Title: US-09-674-266A-181
Perfect score: 1063
Sequence: 1 RLSCAGTLSGSGPHPSRRLT.....DTALLLDNMKKALKLLKTEL 206

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	448	42.1	183	1	XAG_XENLA	P55868 xenopus lae
2	437.5	41.2	185	1	NP77_XENLA	P55869 xenopus lae
3	87.5	8.2	458	1	YM51_YEAST	P42933 saccharomyc
4	85.5	8.0	1829	1	MY5A_CHICK	Q02440 gallus gall
5	84	7.9	811	1	RFXL_YEAST	P48743 saccharomyc
6	83.5	7.9	217	1	YD67_SCHPO	Q10319 schizosacch
7	83	7.8	507	1	HSK1_SCHPO	P50582 schizosacch
8	81.5	7.7	901	1	OCRL_HUMAN	Q01968 homo sapien
9	81.5	7.7	1398	1	TOP2_PLAFK	P41001 plasmodium
10	80	7.5	339	1	YTH6_RHOER	P43484 rhodococcus
11	80	7.5	843	1	YMS1_YEAST	Q05050 saccharomyc
12	79.5	7.5	301	1	ERA_HELPJ	Q9zlw0 helicobacte
13	79	7.4	510	1	CBP1_ORYSA	P37890 oryza sativ
14	78.5	7.4	280	1	AROE_BACSU	P54374 bacillus su
15	78.5	7.4	1262	1	TPP2_MOUSE	Q64514 mus musculu
16	78.5	7.4	1828	1	MY5A_RAT	Q9qyf3 rattus norv
17	78	7.3	337	1	HAIR_DROME	P14003 drosophila
18	78	7.3	887	1	ITH3_RAT	Q63416 rattus norv
19	78	7.3	1853	1	MY5A_MOUSE	Q99104 mus musculu
20	78	7.3	1855	1	MY5A_HUMAN	Q9y4i1 homo sapien
21	76.5	7.2	1025	1	PO21_NASVI	Q03278 nasonia vit
22	76	7.1	543	1	OPPA_ECOLI	P23843 escherichia
23	76	7.1	2145	1	U520_CAEEL	Q9u2g0 caenorhabdi
24	75.5	7.1	580	1	TRE2_MYCTU	Q10769 mycobacteri
25	75.5	7.1	2318	1	NTC3_MOUSE	Q61982 mus musculu
26	75	7.1	1694	1	SN_MOUSE	Q62230 mus musculu
27	74.5	7.0	905	1	ZO3_MOUSE	Q9qxy1 mus musculu
28	74	7.0	260	1	TATD_ECOLI	P27859 escherichia
29	74	7.0	986	1	GM13_RAT	Q62839 rattus norv
30	74	7.0	1034	1	POL_HV2CA	P24107 human immun
31	74	7.0	1651	1	VIT6_CAEEL	P18948 caenorhabdi
32	73.5	6.9	233	1	YEG7_YEAST	P39983 saccharomyc
33	73.5	6.9	695	1	TKT_PICST	P34736 pichia stip
34	73.5	6.9	722	1	YP65_CAEEL	Q09214 caenorhabdi
35	73.5	6.9	971	1	CSE1_PAGMA	Q9ptu3 pagrus majo
36	73	6.9	448	1	GUN_CLOSA	P15704 clostridium
37	73	6.9	522	1	UNC7_CAEEL	Q03412 caenorhabdi
38	72.5	6.8	188	1	KTHY_METJA	Q57741 methanococc
39	72.5	6.8	689	1	YYAL_BACSU	P37512 bacillus su
40	72.5	6.8	823	1	DMSA_RHOCA	Q52675 rhodobacter
41	72	6.8	353	1	UL95_HSV6U	P24444 human herpe
42	72	6.8	443	1	YKJ0_YEAST	P36075 saccharomyc
43	72	6.8	523	1	YQE4_CAEEL	Q17529 caenorhabdi
44	72	6.8	688	1	YP22_YEAST	Q08968 saccharomyc
45	72	6.8	898	1	ZO3_CANFA	O62683 canis famil

ALIGNMENTS

RESULT 1
XAG_XENLA
ID XAG_XENLA STANDARD; PRT; 183 AA.
AC P55868;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Putative secreted protein XAG precursor.
GN XAG.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cement gland;
RX MEDLINE=97003224; PubMed=8850563;
RA Sive H., Bradley L.;
RT "A sticky problem: the Xenopus cement gland as a paradigm for
RT anteroposterior patterning.";
RL Dev. Dyn. 205:265-280(1996).
CC -!- SUBCELLULAR LOCATION: Secreted (Probable).
CC -!- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVEL IN CEMENT GLAND.
CC -!- SIMILARITY: HIGH, TO XENOPUS NP77.
CC ----------
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ----------
DR EMBL; U76752; AAB18819.1; -.
KW Signal.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 183 PUTATIVE SECRETED PROTEIN XAG.
SQ SEQUENCE 183 AA; 20475 MW; 95E12B06FCB8DCE8 CRC64;

Query Match 42.1%; Score 448; DB 1; Length 183;
Best Local Similarity 49.4%; Pred. No. 1e-34;
Matches 88; Conservative 29; Mismatches 51; Indels 10; Gaps 3;

```
Qy      39 AFLLLVALSYTLARDTTVKP---GAKKDTKDSRPKLP------QTLSRGWGDQLIWTQTY 89
           : : || |   |       ||       ||| :   |       :|| |||| : | |||
Db       6 SLVCLVLLCSALGEAVLKKPKKQAGTTDTKTDQEPAPIKTKGLKTLDRGWGESIEWVQTY 65

Qy      90 EEALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLA-EQFVLLNLVYETTDKH 148
             || | |:: :|||||:||||::||:| |||| |  ::  |||| | |::||||:   |::
Db      66 EEGLAKARENNKPLMVIHHLEDCPYSIALKKAFVADRMAQKLAQEDFIMLNLVHPVADEN 125

Qy     149 LSPDGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 206
             |||| ||||::|:|||||||:|: ||| |::|||:  |   |: |||||   |||||
Db     126 QSPDGHYVPRVIFIDPSLTVRSDLKGRYGNKMYAYDADDIPELITNMKKAKSFLKTEL 183
```

```
RESULT 2
NP77_XENLA
ID   NP77_XENLA     STANDARD;      PRT;   185 AA.
AC   P55869;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Putative secreted protein NP77 precursor.
GN   NP77.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Cement gland;
RA   Aberger F., Schueren C., Lepperdinger G., Richter K., Grunz H.;
RL   Submitted (MAR-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Secreted (Probable).
CC   -!- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVEL IN CEMENT GLAND
CC       (POTENTIAL).
CC   -!- SIMILARITY: HIGH, TO XENOPUS XAG.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U82110; AAB49974.1; -.
KW   Signal.
FT   SIGNAL        1     18       POTENTIAL.
FT   CHAIN        19    185       PUTATIVE SECRETED PROTEIN NP77.
SQ   SEQUENCE   185 AA;  20442 MW;  AE3807C926044509 CRC64;

  Query Match             41.2%;  Score 437.5;  DB 1;  Length 185;
  Best Local Similarity   53.5%;  Pred. No. 9.7e-34;
  Matches   84;  Conservative   26;  Mismatches   46;  Indels    1;  Gaps    1;

Qy      51 ARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEEALYKSKTSNKPLMIIHHLD 110
           |  |     || :    : |  :|| |||| : | ||||| | |:: :|||||:||||:
Db      29 AAATDTNGAAKSEPAPVKTKGLKTLDRGWGEDIEWAQTYEEGLAKARENNKPLMVIHHLE 88

Qy     111 ECPHSQALKKVFAENKEIQKLA-EQFVLLNLVYETTDKHLSPDGQYVPRIMFVDPSLTVR 169
           :||:| |||| |  :|  |||| | |::|||:   |:: |||| |||: :|:|||||||
Db      89 DCPYSIALKKAFVADKMAQKLAQEDFIMLNLVHPVADENQSPDGHYVPKGIFIDPSLTVR 148

Qy     170 ADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 206
           :|: ||| |:|||:  |   |: |||||   |||||
Db     149 SDLKGRYGNKLYAYDADDIPELITNMKKAKSFLKTEL 185


RESULT 3
YM51_YEAST
ID   YM51_YEAST     STANDARD;      PRT;   458 AA.
AC   P42933;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Hypothetical 52.2 kDa protein in SGS1-MRPL24 intergenic region.
GN   YMR191W OR YM9646.03.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Pearson D., Bowman S., Barrell B.G., Rajandream M.A.;
RL   Submitted (JAN-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 1-249 FROM N.A.
RC   STRAIN=W303;
RX   MEDLINE=95059068; PubMed=7969174;
RA   Gangloff S., McDonald J.P., Bendixen C., Arthur L., Rothstein R.;
RT   "The yeast type I topoisomerase Top3 interacts with Sgs1, a DNA
RT   helicase homolog: a potential eukaryotic reverse gyrase.";
RL   Mol. Cell. Biol. 14:8391-8398(1994).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z47815; CAA87812.1; -.
DR   EMBL; U22341; AAB60288.1; -.
DR   SGD; S0004803; YMR191W.
KW   Hypothetical protein.
FT   CONFLICT    128    128       N -> Y (IN REF. 2).
FT   CONFLICT    132    135       FVTW -> LALG (IN REF. 2).
SQ   SEQUENCE   458 AA;  52183 MW;  C31107D7B28DAE33 CRC64;

  Query Match             8.2%;  Score 87.5;  DB 1;  Length 458;
  Best Local Similarity   20.1%;  Pred. No. 1.2;
  Matches   55;  Conservative   35;  Mismatches   74;  Indels  109;  Gaps   13;

Qy      15 PSRRLTQG-----RWVRKSRVAMEKIPVSAFLLLVALS---------------------- 47
           |:| :  |       |:| ||| : :|  |    |::||
Db      82 PTRNMAVGGNNWSMWLRMSRVHLRQITKSLDRTLISLSHGNFSHQYNRNIFVTWWKSLFE 141

Qy      48 ----YTLARDTTVKPGAKKDTK-------------DSRPKLPQTLSRG----W----GD 81
               :  |   || |  ::                  | |::|:   ||    |
Db     142 ASTAFRRASGLTVSPLTRRGIARFDHFRPVPNVSKFASFPRVPKGAPRGLFTNWNMTTSK 201

Qy      82 QLIWTQTYEEALYK---SKTSNKPLMI------IHHLDECPHSQ---------------- 116
           :|:  : |  :  |      :|  : :      :  |::| ||
Db     202 RLLGQRAYSTSSIKFTQEAVNNMTISLRCFFNSLGGLNQCSHSNSCKAYQNASNVTSKQD 261

Qy     117 -----ALKKV------FAENKEIQKLAEQFVLLNLVYETTDKHLSPDGQYV--------- 156
               ||||:       |  | |: |: :       :| ||:  ::   | |:
Db     262 HVQPVALKKLSQKDINFIRNLELFKIMK--TQNEVVDETSAYYMEKPGSYIEFTISEFNV 319

Qy     157 -----PRIMFVDPSLTVRADITGRYSNRLYAYE 184
                 : |:||||   ||:     |  | ||
Db     320 NGTFSAPLSFLDPSLL--ADLDEMIRN--YKYE 348


RESULT 4
MY5A_CHICK
ID   MY5A_CHICK     STANDARD;      PRT;  1829 AA.
AC   Q02440;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Myosin Va (Myosin 5A) (Dilute myosin heavy chain, non-muscle) (Myosin
DE   heavy chain P190) (Myosin-V).
GN   MYO5A.
OS , Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=93012002; PubMed=1383040;
RA   Sanders G., Lichte B., Meyer H.E., Kilimann M.W.;
RT   "cDNA encoding the chicken ortholog of the mouse dilute gene product.
RT   Sequence comparison reveals a myosin I subfamily with conserved C-
```

```
RT   terminal domains.";
RL   FEBS Lett. 311:295-298(1992).
RN   [2]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1154-1163.
RC   TISSUE=Brain;
RX   MEDLINE=93107155; PubMed=1469047;
RA   Espreafico E.M., Cheney R.E., Matteoli M., Nascimento A.A.,
RA   de Camilli P.V., Larson R.E., Mooseker M.S.;
RT   "Primary structure and cellular localization of chicken brain
RT   myosin-V (p190), an unconventional myosin with calmodulin light
RT   chains.";
RL   J. Cell Biol. 119:1541-1557(1992).
CC   -!- FUNCTION: PROCESSIVE ACTIN-BASED MOTOR THAT CAN MOVE IN LARGE
CC       STEPS APPROXIMATING THE 36-NM PSEUDO-REPEAT OF THE ACTIN FILAMENT.
CC       MAY BE INVOLVED IN MELANOSOME TRANSPORT, OR ALTERNATIVELY, IT MAY
CC       BE REQUIRED FOR SOME POLARIZATION PROCESS INVOLVED IN DENDRITE
CC       FORMATION (BY SIMILARITY).
CC   -!- SUBUNIT: MAY BE A HOMODIMER, WHICH ASSOCIATES WITH MULTIPLE
CC       CALMODULIN OR MYOSIN LIGHT CHAINS.
CC   -!- SUBCELLULAR LOCATION: GOLGI-DERIVED CYTOPLASMIC MEMBRANES
CC       (POTENTIAL).
CC   -!- TISSUE SPECIFICITY: NEURONAL AND NONNEURONAL CELLS OF THE BRAIN.
CC   -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC   -!- SIMILARITY: CONTAINS 6 IQ DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 DILUTE DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X67251; CAA47673.1; -.
DR   EMBL; Z11718; CAA77782.1; -.
DR   HSSP; P08799; 1MMG.
DR   InterPro; IPR002710; DIL.
DR   InterPro; IPR000048; IQ.
DR   InterPro; IPR001609; myosin_head.
DR   Pfam; PF01843; DIL; 1.
DR   Pfam; PF00612; IQ; 6.
DR   Pfam; PF00063; myosin_head; 1.
DR   PRINTS; PR00193; MYOSINHEAVY.
DR   ProDom; PD000355; myosin_head; 1.
DR   ProDom; PD003376; DIL; 1.
DR   SMART; SM00015; IQ; 6.
DR   SMART; SM00242; MYSc; 1.
DR   PROSITE; PS50096; IQ; 6.
KW   Myosin; Repeat; ATP-binding; Calmodulin-binding; Actin-binding;
KW   Coiled coil; Phosphorylation.
FT   DOMAIN        1    766       MYOSIN HEAD-LIKE.
FT   DOMAIN      767    789       IQ 1.
FT   DOMAIN      790    814       IQ 2.
FT   DOMAIN      815    837       IQ 3.
FT   DOMAIN      838    862       IQ 4.
FT   DOMAIN      863    887       IQ 5.
FT   DOMAIN      888    915       IQ 6.
FT   DOMAIN      916   1239       COILED COIL (POTENTIAL).
FT   DOMAIN     1315   1419       COILED COIL (POTENTIAL).
FT   DOMAIN     1661   1766       DILUTE.
FT   NP_BIND     163    170       ATP (BY SIMILARITY).
FT   DOMAIN      644    666       ACTIN-BINDING (POTENTIAL).
FT   MOD_RES    1734   1734       PHOSPHORYLATION (POTENTIAL).
FT   CONFLICT   1142   1142       E -> EQ (IN REF. 2).
SQ   SEQUENCE   1829 AA;  212381 MW;  0538B278DFC09F6E CRC64;

  Query Match             8.0%;  Score 85.5;  DB 1;  Length 1829;
  Best Local Similarity   28.1%;  Pred. No. 11;
  Matches   27;  Conservative   21;  Mismatches   41;  Indels    7;  Gaps    3;

Qy      63 DTKDSRPKLPQTLSRGWGDQLIWTQTYEEALY-KSKTSNKPLMIIHHLDECPH--SQALK 119
           |  |   |:|:     |  :|  |   : ||: | : |||  :| |  |:  :     |:

Db     507 DLLDEECKMPKGSDDTWAQKLYNTHLNKCALFEKPRLSNKAFIIKHFADKVEYQCEGFLE 566

Qy     120 K----VFAENKEIQKLAEQFVLLNLVYETTDKHLSP 151
           |    |: |  :: | :::| ||  :::  :| :||
Db     567 KNKDTVYEEQIKVLKSSKKFKLLPELFQDEEKAISP 602


RESULT 5
RFXL_YEAST
ID   RFXL_YEAST     STANDARD;      PRT;   811 AA.
AC   P48743;
DT   01-FEB-1996 (Rel. 33, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical 90.6 kDa protein in CBF5-DKA1 intergenic region.
GN   YLR176C OR L9470.18.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA   Favello A., Fulton L., Gattung S., Greco T., Kirsten J.,
RA   Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
RA   Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
RA   Mardis E., Menezes S., Miller N., Nhan M., Pauley A., Peluso D.,
RA   Rifken L., Riles L., Taich A., Trevaskis E., Vignati D.,
RA   Wilcox L., Wohldman P., Vaudin M., Wilson R., Waterston R.;
RL   Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE RFX FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U17246; AAB67470.1; -.
DR   SGD; S0004166; YLR176C.
DR   InterPro; IPR003150; RFX_DNA_binding.
DR   Pfam; PF02257; RFX_DNA_binding; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   811 AA;  90583 MW;  116A88B7DDE4FBF0 CRC64;

  Query Match             7.9%;  Score 84;  DB 1;  Length 811;
  Best Local Similarity   25.7%;  Pred. No. 5.4;
  Matches   47;  Conservative   29;  Mismatches   67;  Indels   40;  Gaps    8;

Qy       8 LSGSGP----------HPSRRLTQGR----WVRKSRVAMEKIPVSAFLLLVALSYTLAR 52
           | ||||           |:: || :     |  : ::: :   ||: |      ||
Db     129 LIGSGPGSHKPSPTPTQPPAQPATQRQPATYSVYPASISLNRSNSSAYPLSFKSEETLNN 188

Qy      53 D--TTVK--------PGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEEALYKSKTSNKP 102
           :  |  |         | : |  | |: |   ::||    ::|  |  |    |||  :
Db     189 NPPTAAKRTNTFPSIPSSTKKQKTSQEKRISSISRRNTQEIIAKQIAEN--NKSKTIEEY 246

Qy     103 LMIIHH-------LDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSPDGQY 155
             |: |       :|    |:|  ::  :|:| ::  : | || |:     :| |    |
Db     247 AQIVKHAEIKVLSMDSQNTSKAALQLAEQNRERER--QVFALLWLMKNCKSQHDS----Y 300

Qy     156 VPR 158
           |||
Db     301 VPR 303


RESULT 6
YD67_SCHPO
ID   YD67_SCHPO     STANDARD;      PRT;   217 AA.
AC   Q10319;
```

```
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Hypothetical 24.9 kDa protein C17G8.07 in chromosome I.
GN   SPAC17G8.07.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   Devlin K., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL   Submitted (FEB-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: TO HUMAN AF9 AND ENL AND YEAST TRANSCRIPTION
CC       INITIATION FACTOR TFIIF SMALL SUBUNIT (TFG3/ANC1).
CC   -!- SIMILARITY: STRONG, TO YEAST YNL107W.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z69795; CAA93690.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   217 AA;  24946 MW;  A274C2AFB0620D7F CRC64;

  Query Match             7.9%;  Score 83.5;  DB 1;  Length 217;
  Best Local Similarity   20.6%;  Pred. No. 1.1;
  Matches   42;  Conservative   33;  Mismatches   80;  Indels   49;  Gaps    8;

Qy      14 HPSRRLTQG-------RWVRKSRVAMEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKD 66
           |  |   :|       :||||            | |          ||   |     | :
Db      39 HTWRIFVEGVDGEDISKWVRK----------VVFKL---------HDTYNNP---TRTIE 76

Qy      67 SRPKLPQTLSRGWGDQLIWTQTYEEALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENK 126
           | |   : :  |||:  |  : :    :  :   | |   |||   |:   :::: |
Db      77 SPPF--EVIETGWGEFDIMVRIF----FAPEAHEKALTFYHHLKLHPYGPRMEEMKASGG 130

Qy     127 EIQKLAEQFVLLNLVYETTDKHLSP----DGQYVPRIMFVDPSLTVRADITGRYSNRLYA 182
           :: :  : :: |  :| | | ||     ||            | | ::    :| :|
Db     131 LVESVQYEEIVFNEPFEYTYKLLSQNPIGDGH----------GLAVESEPDHPFSQQLEQ 180

Qy     183 YEPADTALLLDNMKKALKLLKTEL 206
           |       :  :|| ::: | ::
Db     181 DEADKLDFAIQEVKKTIEMYKQQV 204


RESULT 7
HSK1_SCHPO
ID   HSK1_SCHPO     STANDARD;      PRT;   507 AA.
AC   P50582; O94678;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Cell cycle protein kinase hsk1 (EC 2.7.1.-).
GN   HSK1 OR SPBC776.12C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A. (LONG FORM).
RC   STRAIN=JY2;
RX   MEDLINE=95347336; PubMed=7621824;
RA   Masai H., Miyake T., Arai K.-I.;
RT   "hsk1+, a Schizosaccharomyces pombe gene related to Saccharomyces
RT   cerevisiae CDC7, is required for chromosomal replication.";
RL   EMBO J. 14:3094-3104(1995).
RN   [2]
RP   SEQUENCE FROM N.A. (LONG AND SHORT FORM).
RC   STRAIN=972;
RA   Lyne M., Rajandream M.A., Barrell B.G., Wedler H., Kutzner M.,
RA   Wambutt R.;
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: REQUIRED FOR CHROMOSOMAL REPLICATION.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC       SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC       CDC7 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D50493; BAA09087.1; -.
DR   EMBL; AL035263; CAA22885.1; -.
DR   EMBL; AL035263; CAB38417.1; -.
DR   HSSP; Q16539; 1WFC.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 3.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Cell cycle; Cell division; Serine/threonine-protein kinase;
KW   Transferase; ATP-binding; Alternative splicing.
FT   DOMAIN       68    433       PROTEIN KINASE.
FT   NP_BIND      74     82       ATP (BY SIMILARITY).
FT   BINDING     129    129       ATP (BY SIMILARITY).
FT   ACT_SITE    216    216       BY SIMILARITY.
FT   VARSPLIC     92    119       MISSING (IN SHORT ISOFORM).
SQ   SEQUENCE   507 AA;  58407 MW;  04970E58218441EE CRC64;

  Query Match             7.8%;  Score 83;  DB 1;  Length 507;
  Best Local Similarity   22.4%;  Pred. No. 3.7;
  Matches   28;  Conservative   24;  Mismatches   55;  Indels   18;  Gaps    5;

Qy      83 LIWTQTYEEALYKSKTSNKP----LMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLL 138
           ::|       ::|| |  :||     : :  |::|       |:: ||    : |   |: |
Db     380 ILWASCGSASIYKEKLRHKPSQEERLCLDFLEKCLELDCNKRISAE----EALDHDFLYL 435

Qy     139 -NLVYETTDKHLSPDGQYVPRIMFVDPSLTVRADITGRYSNRLYAY---EPADTALLLDN 194
            || ||  |   : |       | : |      |:| :: : :  :    |  |    |
Db     436 DNLAYEKKDDDTAFDNS------FGETSFEKDEDLTAKHLSHILDFKEQEETDEPTSLSK 489

Qy     195 MKKAL 199
           |:::
Db     490 RKRSI 494


RESULT 8
OCRL_HUMAN
ID   OCRL_HUMAN     STANDARD;      PRT;   901 AA.
AC   Q01968; Q15684; O60800; Q9UJG5; Q15774; Q9UMA5;
DT   01-JUL-1993 (Rel. 26, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Inositol polyphosphate 5-phosphatase OCRL-1 (EC 3.1.3.-) (Lowe's
DE   oculocerebrorenal syndrome protein).
GN   OCRL OR OCRL1 OR INPP5F.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM B).
```

```
RC   TISSUE=Kidney;
RX   MEDLINE=92334430; PubMed=1321346;
RA   Attree O., Olivos I.M., Okabe I., Bailey L.C., Nelson D.L.,
RA   Lewis R.A., McInnes R.R., Nussbaum R.L.;
RT   "The Lowe's oculocerebrorenal syndrome gene encodes a protein highly
RT   homologous to inositol polyphosphate-5-phosphatase.";
RL   Nature 358:239-242(1992).
RN   [2]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC   TISSUE=Brain;
RX   MEDLINE=97201100; PubMed=9048911;
RA   Nussbaum R.L., Orrison B.M., Janne P.A., Charnas L., Chinault A.C.;
RT   "Physical mapping and genomic structure of the Lowe syndrome gene
RT   OCRL1.";
RL   Hum. Genet. 99:145-150(1997).
RN   [3]
RP   SEQUENCE OF 81-901 FROM N.A. (ISOFORMS A AND B).
RA   Pavitt R.;
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 426-901 FROM N.A. (ISOFORM B).
RA   Mistry S.;
RL   Submitted (MAY-1995) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE OF 814-843 FROM N.A.
RX   MEDLINE=93278398; PubMed=8504307;
RA   Leahey A.M., Charnas L.R., Nussbaum R.L.;
RT   "Nonsense mutations in the OCRL-1 gene in patients with the
RT   oculocerebrorenal syndrome of Lowe.";
RL   Hum. Mol. Genet. 2:461-463(1993).
RN   [6]
RP   CHARACTERIZATION.
RX   MEDLINE=95281554; PubMed=7761412;
RA   Zhang X., Jefferson A.B., Auethavekiat V., Majerus P.W.;
RT   "The protein deficient in Lowe syndrome is a
RT   phosphatidylinositol-4,5-bisphosphate 5-phosphatase.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:4853-4856(1995).
RN   [7]
RP   CHARACTERIZATION.
RX   MEDLINE=98104142; PubMed=9430698;
RA   Zhang X., Hartz P.A., Philip E., Racusen L.C., Majerus P.W.;
RT   "Cell lines from kidney proximal tubules of a patient with Lowe
RT   syndrome lack OCRL inositol polyphosphate 5-phosphatase and
RT   accumulate phosphatidylinositol 4,5-bisphosphate.";
RL   J. Biol. Chem. 273:1574-1582(1998).
RN   [8]
RP   VARIANTS LS T-367 DEL; G-451; S-463 AND R-524.
RX   MEDLINE=97342912; PubMed=9199559;
RA   Lin T., Orrison B.M., Leahey A.-M., Suchy S.F., Bernard D.J.,
RA   Lewis R.A., Nussbaum R.L.;
RT   "Spectrum of mutations in the OCRL1 gene in the Lowe oculocerebrorenal
RT   syndrome.";
RL   Am. J. Hum. Genet. 60:1384-1388(1997).
RN   [9]
RP   VARIANTS LS TYR-375; GLN-500; ASP-508 AND CYS-513.
RX   MEDLINE=98347141; PubMed=9682219;
RA   Lin T., Orrison B.M., Suchy S.F., Lewis R.A., Nussbaum R.L.;
RT   "Mutations are not uniformly distributed throughout the OCRL1 gene in
RT   Lowe syndrome patients.";
RL   Mol. Genet. Metab. 64:58-61(1998).
RN   [10]
RP   VARIANTS LS GLN-500 AND GLN-524.
RX   MEDLINE=98293952; PubMed=9632163;
RA   Kawano T., Indo Y., Nakazato H., Shimadzu M., Matsuda I.;
RT   "Oculocerebrorenal syndrome of Lowe: three mutations in the OCRL1 gene
RT   derived from three patients with different phenotypes.";
RL   Am. J. Med. Genet. 77:348-355(1998).
RN   [11]
RP   VARIANT LS ARG-522.
RX   MEDLINE=99002770; PubMed=9788721;
RA   Kubota T., Sakurai A., Arakawa K., Shimazu M., Wakui K., Furihata K.,
RA   Fukushima Y.;
RT   " Identification of two novel mutations in the OCRL1 gene in Japanese
RT   families with Lowe syndrome.";
RL   Clin. Genet. 54:199-202(1998).
RN   [12]
RP   VARIANTS LS GLU-357; GLU-421; ASP-424 AND TYR-498.
RX   MEDLINE=20382695; PubMed=10923037;
RA   Monnier N., Satre V., Lerouge E., Berthoin F., Lunardi J.;
RT   "OCRL1 mutation analysis in French Lowe syndrome patients:
RT   implications for molecular diagnosis strategy and genetic
RT   counseling.";
RL   Hum. Mutat. 16:157-165(2000).
RN   [13]
RP   VARIANTS LS PRO-337 AND GLY-372.
RX   MEDLINE=20232168; PubMed=10767176;
RA   Roschinger W., Muntau A.C., Rudolph G., Roscher A.A.,
RA   Kammerer S.;
RT   "Carrier assessment in families with Lowe oculocerebrorenal syndrome:
RT   novel mutations in the OCRL1 gene and correlation of direct DNA
RT   diagnosis with ocular examination.";
RL   Mol. Genet. Metab. 69:213-222(2000).
CC   -!- FUNCTION: CONVERTS PHOSPHATIDYLINOSITOL 4,5-BISPHOSPHATE TO
CC       PHOSPHATIDYLINOSITOL 4-PHOSPHATE. ALSO CONVERTS INOSITOL 1,4,5-
CC       TRISPHOSPHATE TO INOSITOL 1,4-BISPHOSPHATE AND INOSITOL 1,3,4,5-
CC       TETRAKISPHOSPHATE TO INOSITOL 1,3,4-TRISPHOSPHATE. MAY FUNCTION IN
CC       LYSOSOMAL MEMBRANE TRAFFICKING BY REGULATING THE SPECIFIC POOL OF
CC       PHOSPHATIDYLINOSITOL 4,5-BISPHOSPHATE THAT IS ASSOCIATED WITH
CC       LYSOSOMES.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A (SHOWN HERE) AND B; ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: BRAIN, SKELETAL MUSCLE, HEART, KIDNEY,
CC       LUNG, PLACENTA, AND FIBROBLASTS.
CC   -!- DISEASE: DEFECTS IN OCRL ARE THE CAUSE OF LOWE'S OCULOCEREBRORENAL
CC       SYNDROME, A DISEASE CHARACTERIZED BY HYDROPHTHALMIA, CATARACT,
CC       MENTAL RETARDATION, VITAMIN D-RESISTANT RICKETS, AMINOACIDURIA,
CC       AND REDUCED AMMONIA PRODUCTION BY THE KIDNEY.
CC   -!- SIMILARITY: BELONGS TO THE INOSITOL-1,4,5-TRISPHOSPHATE
CC       5-PHOSPHATASE TYPE II FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 GAP DOMAIN.
CC   -!- DATABASE: NAME=Lowe Syndrome mutation database;
CC       WWW="http://www.nhgri.nih.gov/DIR/GDRB/Lowe/ocrl1_mut_db.html".
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M88162; AAA59964.1; ALT_INIT.
DR   EMBL; U57627; AAB03839.2; -.
DR   EMBL; AL022162; CAA18150.1; -.
DR   EMBL; AL022162; CAA18151.1; -.
DR   EMBL; Z73496; CAA97842.1; -.
DR   EMBL; S62085; AAB26926.1; -.
DR   MIM; 309000; -.
DR   InterPro; IPR000300; IPPc.
DR   InterPro; IPR000198; RhoGAP.
DR   Pfam; PF00783; IPPc; 1.
DR   Pfam; PF00620; RhoGAP; 1.
DR   SMART; SM00128; IPPc; 1.
DR   SMART; SM00324; RhoGAP; 1.
KW   Hydrolase; Alternative splicing; Disease mutation.
FT   DOMAIN      735    875       GAP DOMAIN.
FT   VARSPLIC    707    714       MISSING (IN ISOFORM B).
FT   VARIANT     337    337       R -> P (IN LOWE SYNDROME).
FT                                /FTId=VAR_010169.
FT   VARIANT     357    357       G -> E (IN LOWE SYNDROME).
FT                                /FTId=VAR_010170.
FT   VARIANT     367    367       MISSING (IN LOWE SYNDROME).
FT                                /FTId=VAR_010171.
FT   VARIANT     372    372       V -> G (IN LOWE SYNDROME).
FT                                /FTId=VAR_010172.
FT   VARIANT     375    375       H -> Y (IN LOWE SYNDROME).
```

```
FT                                   /FTId=VAR_010173.
FT   VARIANT     421    421          G -> E (IN LOWE SYNDROME).
FT                                   /FTId=VAR_010174.
FT   VARIANT     424    424          N -> D (IN LOWE SYNDROME).
FT                                   /FTId=VAR_010175.
FT   VARIANT     451    451          D -> G (IN LOWE SYNDROME).
FT                                   /FTId=VAR_010176.
FT   VARIANT     463    463          F -> S (IN LOWE SYNDROME).
FT                                   /FTId=VAR_010177.
FT   VARIANT     498    498          C -> Y (IN LOWE SYNDROME).
FT                                   /FTId=VAR_010178.
FT   VARIANT     500    500          R -> G (IN LOWE SYNDROME).
FT                                   /FTId=VAR_010179.
FT   VARIANT     500    500          R -> Q (IN LOWE SYNDROME).
FT                                   /FTId=VAR_010180.
FT   VARIANT     508    508          V -> D (IN LOWE SYNDROME).
FT                                   /FTId=VAR_010181.
FT   VARIANT     513    513          Y -> C (IN LOWE SYNDROME).
FT                                   /FTId=VAR_010182.
FT   VARIANT     522    522          S -> R (IN LOWE SYNDROME).
FT                                   /FTId=VAR_010183.
FT   VARIANT     524    524          H -> Q (IN LOWE SYNDROME).
FT                                   /FTId=VAR_010184.
FT   VARIANT     524    524          H -> R (IN LOWE SYNDROME).
FT                                   /FTId=VAR_010185.
FT   VARIANT     526    526          P -> F (IN LOWE SYNDROME).
FT                                   /FTId=VAR_010186.
FT   VARIANT     533    533          I -> S (IN LOWE SYNDROME).
FT                                   /FTId=VAR_010187.
FT   VARIANT     768    768          I -> N (IN LOWE SYNDROME).
FT                                   /FTId=VAR_010188.
FT   VARIANT     797    797          A -> P (IN LOWE SYNDROME).
FT                                   /FTId=VAR_010189.
FT   CONFLICT    585    585          G -> E (IN REF. 3 AND 4).
SQ   SEQUENCE   901 AA;  104132 MW;  90B6F6CDDD354BFE CRC64;

  Query Match             7.7%;  Score 81.5;  DB 1;  Length 901;
  Best Local Similarity   23.7%;  Pred. No. 11;
  Matches   45;  Conservative   32;  Mismatches   46;  Indels   67;  Gaps   13;

Qy      55 TVKPGAKKDTKDSRPK--LPQTLSRGWGDQLIWTQTYEEAL-YKS----KTS-NKPLMII 106
           | |  :| |  ||  |  :|      | |:::|  |    | |:|    ||| :||:  :
Db     476 TYKYDSKTDRWDSSGKCRVP-----AWCDRILWRGTNVNQLNYRSHMELKTSDHKPVSAL 530

Qy     107 HHL-----DECPHSQALKKVFAENKEIQKLAE-----------QFVLLNLVYETTDK-- 147
            |:     ||    :  :||| ::  |    |            :||  |: :    |
Db     531 FHIGVKVVDE----RRYRKVFEDSVRIMDRMENDFLPSLELSRREFVFENVKFRQLQKGK 586

Qy     148 -HLSPDGQ------YVPRIM---------------FVDPSLTVRADITGRYSNRLYAYE 184
            :| :||      ::|::                :::|: ||  ||:       |  |
Db     587 FQISNNGQVPCHFSFIPKLNDSQYCKPWLRAEPFEGYLEPNETV--DIS------LDVYV 638

Qy     185 PADTALLLDN 194
             |:  :|::
Db     639 SKDSVTILNS 648


RESULT 9
TOP2_PLAFK
ID   TOP2_PLAFK     STANDARD;      PRT;  1398 AA.
AC   P41001;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   DNA topoisomerase II (EC 5.99.1.3).
GN   TOP2.
OS   Plasmodium falciparum (isolate K1 / Thailand).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5839;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94316496; PubMed=8041616;
RA   Cheesman S., McAleese S., Goman M., Johnson D., Horrocks P.,
RA   Ridley R.G., Kilbey B.J.;
RT   "The gene encoding topoisomerase II from Plasmodium falciparum.";
RL   Nucleic Acids Res. 22:2547-2551(1994).
CC   -!- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT
CC       BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II
CC       MAKES DOUBLE-STRAND BREAKS.
CC   -!- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining
CC       of double-stranded DNA.
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH
CC       NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES
CC       RELAX ONLY NEGATIVE SUPERCOILS.
CC   -!- SIMILARITY: BELONGS TO THE TYPE II TOPOISOMERASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X79345; -; NOT_ANNOTATED_CDS.
DR   HSSP; P06786; 1BGW.
DR   InterPro; IPR003957; CBFA_NFYB_topis.
DR   InterPro; IPR001241; DNA_topoisoII.
DR   InterPro; IPR002205; DNA_topoisoIV.
DR   InterPro; IPR003594; HATPase_c.
DR   Pfam; PF00204; DNA_topoisoII; 1.
DR   Pfam; PF00521; DNA_topoisoIV; 1.
DR   Pfam; PF02518; HATPase_c; 1.
DR   PRINTS; PR00615; CCAATSUBUNTA.
DR   PRINTS; PR00418; TPI2FAMILY.
DR   ProDom; PD000616; DNA_topoisoII; 1.
DR   SMART; SM00433; TOP2c; 1.
DR   SMART; SM00434; TOP4c; 1.
DR   PROSITE; PS00177; TOPOISOMERASE_II; 1.
KW   Isomerase; Topoisomerase; DNA-binding; ATP-binding; Nuclear protein.
FT   NP_BIND     144    149       ATP (POTENTIAL).
FT   ACT_SITE    830    830       DNA CLEAVAGE (BY SIMILARITY).
FT   DOMAIN      271    281       POLY-ASN.
FT   DOMAIN      308    316       POLY-ASN.
FT   DOMAIN     1089   1093       POLY-LYS.
FT   DOMAIN     1227   1234       POLY-LYS.
SQ   SEQUENCE   1398 AA;  161029 MW;  BAAD7BEE88FE5BE9 CRC64;

  Query Match             7.7%;  Score 81.5;  DB 1;  Length 1398;
  Best Local Similarity   21.9%;  Pred. No. 18;
  Matches   30;  Conservative   25;  Mismatches   47;  Indels   35;  Gaps    5;

Qy      85 WTQTYEEALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQ-----FVLLN 139
           ||| |:| | :  |  |  :|: ::|   |      :  :  ::||  |:     | | :
Db     956 WTQDYKEFLEELLTDEKHQLILDYIDNSSHEDICFTIKMDPAKLQKAEEEGLEKVFKLKS 1015

Qy     140 LVYETTDKHLSPDGQYVPRIMFVDPSLTVRADITGRYS----------NRLYAYEPADT 188
             :  |              :   ||:| ::     |||             || |||
Db    1016 TLTTT-------------NMTLFDPNLKLQ-----RYSTELDILKEFCYQRLKAYENR-K 1056

Qy     189 ALLLDNMKKALKLLKTE 205
           : |:  ::|  :::  :
Db    1057 SYLISKLEKEKRIISNK 1073


RESULT 10
YTH6_RHOER
ID   YTH6_RHOER     STANDARD;      PRT;   339 AA.
AC   P43484;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical 37.5 kDa protein in thcR 5'region (ORF6).
OS   Rhodococcus erythropolis.
```

```
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.
OX   NCBI_TaxID=1833;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NI86/21;
RX   MEDLINE=96082876; PubMed=7583123;
RA   Tamura T., Nagy I., Lupas A., Lottspeich F., Cejka Z., Schoofs G.,
RA   Tanaka K., de Mot R., Baumeister W.;
RT   "The first characterization of a eubacterial proteasome: the 20S
RT   complex of Rhodococcus.";
RL   Curr. Biol. 5:766-774(1995).
RN   [2]
RP   SEQUENCE OF 1-333 FROM N.A.
RC   STRAIN=NI86/21;
RX   MEDLINE=95138028; PubMed=7836301;
RA   Nagy I., Schoofs G., Compernolle F., Proost P., Vanderleyden J.,
RA   de Mot R.;
RT   "Degradation of the thiocarbamate herbicide EPTC (S-ethyl
RT   dipropylcarbamothioate) and biosafening by Rhodococcus sp. strain
RT   NI86/21 involve an inducible cytochrome P-450 system and aldehyde
RT   dehydrogenase.";
RL   J. Bacteriol. 177:676-687(1995).
CC   -!- SIMILARITY: STRONG, TO M.TUBERCULOSIS RV2097C AND M.LEPRAE
CC       MLCB2533.24.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U26421; AAC45738.1; -.
DR   EMBL; U17130; AAC45747.1; -.
DR   InterPro; IPR004347; DUF245.
DR   Pfam; PF03136; DUF245; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   339 AA;  37462 MW;  23CB953F88739599 CRC64;

  Query Match             7.5%;  Score 80;  DB 1;  Length 339;
  Best Local Similarity   29.5%;  Pred. No. 4.2;
  Matches   31;  Conservative   20;  Mismatches   36;  Indels   18;  Gaps    5;

Qy     101 KPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSPDGQYVPRIM 160
           :|:  :||:   |  :|     |: :|:  || | : |: |    || :| :|   ||:
Db     122 RPVTAVHHISHDPTLRA-TVALADGRELTGLALQRIYLDRV----DKFMSAEGNDDPRV- 175

Qy     161 FVDPSLTVRADITGRYSNRLYAYE--PADTALLLDNMKKALKLLK 203
                 ||:  :::  |   |  | : | |||   | |:||:
Db     176 ---------ADLLEKWAMVLDLLERDPMECAHLLDWPAK-LRLLE 210


RESULT 11
YMS1_YEAST
ID   YMS1_YEAST     STANDARD;      PRT;   843 AA.
AC   Q05050;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Hypothetical 93.3 kDa protein in TAP42-CYK2 intergenic region.
GN   YMR031C OR YM9973.04C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Bowman S., Barrell B.G., Rajandream M.A.;
RL   Submitted (MAY-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: TO YEAST YKL050C.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z49213; CAA89146.1; -.
DR   SGD; S0004633; YMR031C.
KW   Hypothetical protein.
SQ   SEQUENCE   843 AA;  93345 MW;  0068D65A229B941D CRC64;

  Query Match             7.5%;  Score 80;  DB 1;  Length 843;
  Best Local Similarity   21.3%;  Pred. No. 13;
  Matches   38;  Conservative   29;  Mismatches   57;  Indels   54;  Gaps    7;

Qy      20 TQGRWVRKSRVAMEKIPVSAFLLLVA-----------LSYTLARDTTVKPGAKKDTKDS 67
           |  |   : : : |||  | |  |||           |  | :::   |   ::: ||
Db     453 TANRHQTEKKTSQEKIKAS-FDALVARMDTKVAERETLLEDTKSKEIEFKKQMQQELKDE 511

Qy      68 RPKLPQTLSRGWGDQLIWTQTYEEALYKSKTSNKPLMIIHHLDECPHSQALKKVFAE--- 124
           : :| | |         | :  |: : :::   : |:  :| |:  :::|  |   |
Db     512 KARLDQDLEE------WGKKCEQDITEARKEQEELLKPYH-DDLANAEAEHKTLVEERD 563

Qy     125 --NKEIQKLAEQFV-------------------------LLNL--VYETTDKHLSPD 152
             | || :| :  |                          || |    |: : ||: |
Db     564 EINAEISRLQDAIVDHKRKISGYGNDLDAQKNRNIREDDKLLELGQTKESLESHLNDD 621


RESULT 12
ERA_HELPJ
ID   ERA_HELPJ      STANDARD;      PRT;   301 AA.
AC   Q9ZLW0;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   GTP-binding protein era homolog.
GN   ERA OR JHP0466.
OS   Helicobacter pylori J99 (Campylobacter pylori J99).
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
OX   NCBI_TaxID=85963;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99120557; PubMed=9923682;
RA   Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA   Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA   Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA   Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA   Trust T.J.;
RT   "Genomic sequence comparison of two unrelated isolates of the human
RT   gastric pathogen Helicobacter pylori.";
RL   Nature 397:176-180(1999).
CC   -!- FUNCTION: BINDS BOTH GDP AND GTP, HAS AN INTRINSIC GTPASE ACTIVITY
CC       AND IS ESSENTIAL FOR CELL GROWTH (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ERA/TRME FAMILY OF GTP-BINDING
CC       PROTEINS. ERA SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 KH DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE001480; AAD06035.1; -.
DR   HSSP; P06616; 1EGA.
DR   InterPro; IPR004087; KH.
DR   InterPro; IPR004044; KH_TYPE_2.
```

```
DR   Pfam; PF00013; KH-domain; 1.
DR   PROSITE; PS50823; KH_TYPE_2; 1.
KW   GTP-binding; RNA-binding; Complete proteome.
FT   NP_BIND      12     19       GTP (POTENTIAL).
FT   NP_BIND      64     68       GTP (POTENTIAL).
FT   NP_BIND     122    125       GTP (POTENTIAL).
FT   DOMAIN      204    280       KH.
SQ   SEQUENCE   301 AA;  34485 MW;  F668C7730882EDBF CRC64;

  Query Match             7.5%;  Score 79.5;  DB 1;  Length 301;
  Best Local Similarity   27.8%;  Pred. No. 4;
  Matches   25;  Conservative   13;  Mismatches   33;  Indels   19;  Gaps    4;

Qy      89 YEEALYKSKTSNKP-LMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVL---------- 137
           ||| |       || :: :  :|   | | |:|:    :| || : ||:
Db     106 YEEFL---SLCQKPHILALSKIDTATHKQVLQKL----QEYQKYSSQFLALVPLSAKKSQ 158

Qy     138 -LNLVYETTDKHLSPDGQYVPRIMFVDPSL 166
            || : |   |||||      : :  |  :
Db     159 NLNALLECISKHLSPSAWLFEKDLMSDEKM 188


RESULT 13
CBP1_ORYSA
ID   CBP1_ORYSA     STANDARD;      PRT;   510 AA.
AC   P37890;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Serine carboxypeptidase I precursor (EC 3.4.16.5) (Carboxypeptidase
DE   C).
GN   CBP1.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. YUKIHIKARI;
RX   MEDLINE=94213891; PubMed=8161571;
RA   Washio K., Ishikawa K.;
RT   "Cloning and sequencing of the gene for type I carboxypeptidase in
RT   rice.";
RL   Biochim. Biophys. Acta 1199:311-314(1994).
CC   -!- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a
CC       broad specificity.
CC   -!- PTM: THREE DISULFIDE BONDS ARE PRESENT (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10; ALSO KNOWN AS THE
CC       SERINE CARBOXYPEPTIDASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D17586; BAA04510.1; -.
DR   PIR; S43516; S43516.
DR   HSSP; P08819; 1WHT.
DR   MEROPS; S10.001; -.
DR   InterPro; IPR000379; Est_lip_thioest_actsite.
DR   InterPro; IPR001563; Serine_carbpept.
DR   Pfam; PF00450; serine_carbpept; 1.
DR   PRINTS; PR00724; CRBOXYPTASEC.
DR   PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
DR   PROSITE; PS00342; MICROBODIES_CTER; 1.
DR   PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.
KW   Hydrolase; Carboxypeptidase; Glycoprotein; Zymogen; Signal.
FT   SIGNAL       1     25       POTENTIAL.
FT   PROPEP      26     36       POTENTIAL.
FT   CHAIN        37    510       SERINE CARBOXYPEPTIDASE I.
FT   ACT_SITE    194    194       BY SIMILARITY.
FT   ACT_SITE    434    434       BY SIMILARITY.
FT   ACT_SITE    487    487       BY SIMILARITY.
FT   CARBOHYD    154    154       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    268    268       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    418    418       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   SITE        508    510       MICROBODY TARGETING SIGNAL (POTENTIAL).
SQ   SEQUENCE   510 AA;  55709 MW;  28896247FA1371CF CRC64;

  Query Match             7.4%;  Score 79;  DB 1;  Length 510;
  Best Local Similarity   37.5%;  Pred. No. 8.7;
  Matches   30;  Conservative    5;  Mismatches   33;  Indels   12;  Gaps    4;

Qy      73 QTLSRG--WGDQLIWTQTYEEALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQK 130
           ||   | |       |   | |||| ||   | |   |: | ||: :|||   |    |
Db     259 QTACHGNYWNTT---TDKCENALYKVDTSINDLNIYDILEPCYHSKTIKKVTPAN---TK 312

Qy     131 LAEQFVLLNLVYETTDKHLS 150
           | : |  |     || | |:
Db     313 LPKSFQHLG----TTTKPLA 328


RESULT 14
AROE_BACSU
ID   AROE_BACSU     STANDARD;      PRT;   280 AA.
AC   P54374;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Shikimate 5-dehydrogenase (EC 1.1.1.25).
GN   AROD.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / JH642;
RA   Kobayashi Y., Mizuno M., Masuda S., Takemaru K., Hosono S.,
RA   Sato T., Takeuchi M.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: Shikimate + NADP(+) = 5-dehydroshikimate +
CC       NADPH.
CC   -!- PATHWAY: FOURTH STEP IN THE BIOSYNTHESIS OF CHORISMATE WITHIN THE
CC       BIOSYNTHESIS OF AROMATIC AMINO ACIDS (THE SHIKIMATE PATHWAY).
CC   -!- SIMILARITY: BELONGS TO THE SHIKIMATE DEHYDROGENASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D84432; BAA12445.1; -.
DR   EMBL; Z99117; CAB14508.1; -.
DR   SubtiList; BG11522; aroD.
DR   InterPro; IPR002907; Shikimate_DH.
DR   Pfam; PF01488; Shikimate_DH; 1.
KW   Aromatic amino acid biosynthesis; Oxidoreductase; NADP;
KW   Complete proteome.
SQ   SEQUENCE   280 AA;  30642 MW;  1EE0CE9C2EDA553F CRC64;

  Query Match             7.4%;  Score 78.5;  DB 1;  Length 280;
  Best Local Similarity   24.8%;  Pred. No. 4.5;
  Matches   39;  Conservative   17;  Mismatches   58;  Indels   43;  Gaps    5;

Qy      80 GDQLIWTQTYEEALYKS--KTSNKPLMIIHHL----------DECPH 114
           ||:|:   |  |   ||  |  :||:  :  |           | |
Db      93 GDKLVGYNTDGEGFVKSLMKVLDKPISELSFLMIGAGGAARAIFTTFVRNTPKKFDICNR 152
```

```
Qy    115 SQALKKVFAE------NKE---IQKLAEQFVLLNLVYETTDKHLSPDGQYVPRIMFVDPS 165
Db    153 TLEKAKRLTEATPSFHNKEVLSIKEAEERLEQYDVIIHTTSVGMYPNVDDVPLSLQRAAS 212

Qy    166 LTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLL 202
Db    213 SAVVCDI---------VYNPIQTALLKEASQKGLKTL 240
```

RESULT 15
TPP2_MOUSE
ID   TPP2_MOUSE   STANDARD;   PRT; 1262 AA.
AC   Q64514;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Tripeptidyl-peptidase II (EC 3.4.14.10) (TPP-II) (Tripeptidyl
DE   aminopeptidase).
GN   TPP2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LEADEN X A1;
RX   MEDLINE=95091686; PubMed=7998988;
RA   Tomkinson B.;
RT   "Characterization of cDNA for murine tripeptidyl-peptidase II reveals
RT   alternative splicing.";
RL   Biochem. J. 304:517-523(1994).
CC   -!- CATALYTIC ACTIVITY: Release of an N-terminal tripeptide from a
CC   polypeptide.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC   SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC   SUBTILASE FAMILY.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; X81323; CAA57103.1; -.
DR   HSSP; P00782; 2SBT.
DR   MEROPS; S08.090; -.
DR   MGD; MGI:102724; Tpp2.
DR   InterPro; IPR000209; Peptidase_S8.
DR   Pfam; PF00082; Peptidase_S8; 3.
DR   PRINTS; PR00723; SUBTILISIN.
DR   PROSITE; PS00136; SUBTILASE_ASP; FALSE_NEG.
DR   PROSITE; PS00137; SUBTILASE_HIS; 1.
DR   PROSITE; PS00138; SUBTILASE_SER; 1.
KW   Hydrolase; Alternative splicing.
FT   ACT_SITE   44   44   CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   264   264   CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   449   449   CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   VARSPLIC   985   997   MISSING (IN SHORT ISOFORM).
SQ   SEQUENCE   1262 AA; 139878 MW; D50D22C85544B034 CRC64;

  Query Match   7.4%; Score 78.5; DB 1; Length 1262;
  Best Local Similarity   22.9%; Pred. No. 31;
  Matches   44; Conservative   31; Mismatches   78; Indels   39; Gaps   7;

```
Qy      5 AGTLSGSGPHPSRRLTQGRWVRKSRVAMEKIPVSAFLLLVALSYTLARDTTVKPGAK--- 61
Db    970 AGSLTLSKTELGKKAGQSAAKRQGKFKKDVIPVHYYL--------IPPPTKIKNGSKDKE 1021

Qy     62 ------KDTKDSRPKLPQTLSRGWGDQLIWTQTYEEALYKSKTSNKPLMI--IHHLDECP 113
Db   1022 KDSEKEKDLKEEFTEALRDLKIQWMTKLDSTDIYNE-LKETYPAYLPLYVARLHQLD--- 1077

Qy    114 HSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSPDGQYVPRIMFVDP---SLTVRA 170
Db   1078 ---------AEKERMKRLNEIVDAANAVI----SHIDQTALAVYIAMKTDPRPDAATIKN 1124

Qy    171 DITGRYSNRLYA 182
Db   1125 DMDKQKSTLIDA 1136
```

Search completed: October 9, 2002, 16:40:31
Job time : 15 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 16:39:22 ; Search time 27 Seconds
(without alignments)
1319.887 Million cell updates/sec

Title: US-09-674-266A-181
Perfect score: 1063
Sequence: 1 RLSCAGTLSGSGPHPSRRLT.....DTALLLDNMKKALKLLKTEL 206

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	899	84.6	175	4	O95994	O95994 homo sapien
2	828	77.9	175	11	O88312	O88312 mus musculu
3	548	51.6	159	13	Q90Y05	Q90y05 xenopus lae
4	447	42.1	185	13	O42251	O42251 xenopus lae
5	227	21.4	172	4	Q96H50	Q96h50 homo sapien
6	224.5	21.1	170	11	Q9CQU0	Q9cqu0 mus musculu
7	222	20.9	172	4	O95881	O95881 homo sapien
8	199	18.7	186	5	Q9NA78	Q9na78 caenorhabdi
9	165	15.5	257	5	Q9N5S7	Q9n5s7 caenorhabdi
10	119	11.2	150	17	O27777	O27777 methanother
11	89	8.4	950	11	Q924S5	Q924s5 rattus norv
12	88.5	8.3	737	2	Q9AF04	Q9af04 frankia sp.
13	87	8.2	348	16	Q9Z6X3	Q9z6x3 chlamydia p
14	84.5	7.9	1151	16	Q929J2	Q929j2 listeria in
15	84	7.9	435	5	Q9BJK8	Q9bjk8 caenorhabdi
16	83.5	7.9	918	10	Q9CA88	Q9ca88 arabidopsis
17	83.5	7.9	1011	5	Q9U9P6	Q9u9p6 drosophila
18	83.5	7.9	1409	5	Q9V3A7	Q9v3a7 drosophila
19	83.5	7.9	1409	5	Q9U6I2	Q9u6i2 drosophila
20	83	7.8	786	4	Q9H747	Q9h747 homo sapien
21	83	7.8	976	5	Q9U158	Q9u158 leishmania
22	82.5	7.8	788	16	O67557	O67557 aquifex aeo
23	81.5	7.7	649	10	Q9C5B1	Q9c5b1 arabidopsis
24	81.5	7.7	1158	3	Q9UTR5	Q9utr5 schizosacch
25	81	7.6	610	5	Q9XVM2	Q9xvm2 caenorhabdi
26	81	7.6	635	5	Q09598	Q09598 caenorhabdi
27	80.5	7.6	875	12	Q9QGG9	Q9qgg9 apple mosai
28	80	7.5	154	16	Q9PCE7	Q9pce7 xylella fas
29	80	7.5	259	10	Q9M9H0	Q9m9h0 arabidopsis
30	80	7.5	623	10	Q9FFR8	Q9ffr8 arabidopsis
31	80	7.5	861	5	Q9NKX3	Q9nkx3 bombyx mori
32	79.5	7.5	277	3	O13703	O13703 schizosacch
33	79.5	7.5	536	10	Q93VL7	Q93vl7 arabidopsis
34	79.5	7.5	642	16	Q9KFD1	Q9kfd1 bacillus ha
35	79.5	7.5	921	10	Q9SKT6	Q9skt6 arabidopsis
36	79	7.4	275	2	O87445	O87445 brachyspira
37	79	7.4	337	5	Q95NH3	Q95nh3 drosophila
38	79	7.4	477	3	O13745	O13745 schizosacch
39	79	7.4	480	5	Q9VMD8	Q9vmd8 drosophila
40	78.5	7.4	302	11	Q922K4	Q922k4 mus musculu
41	78.5	7.4	339	6	Q95LR8	Q951r8 macaca fasc
42	78.5	7.4	917	10	O49498	O49498 arabidopsis
43	78.5	7.4	1033	10	Q9CAA7	Q9caa7 arabidopsis
44	78	7.3	127	11	Q9CTG5	Q9ctg5 mus musculu
45	78	7.3	337	5	Q95UA9	Q95ua9 drosophila

ALIGNMENTS

RESULT 1
O95994
ID O95994 PRELIMINARY; PRT; 175 AA.
AC O95994;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SECRETED CEMENT GLAND PROTEIN XAG-2 HOMOLOG (ANTERIOR GRADIENT 2
DE (XENEPUS LAEVIS) HOMOLOG).
GN HAG-2/R OR HAG-2/C.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BREAST;
RX MEDLINE=99009231; PubMed=9790916;
RA Thompson D.A., Weigel R.J.;
RT "hAG-2, the human homologue of the Xenopus laevis cement gland gene
RT XAG-2, is coexpressed with estrogen receptor in breast cancer cell
RT lines.";
RL Biochem. Biophys. Res. Commun. 251:111-116(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Zhang J.S., Smith D.I.;
RT "Human homolog of XAG is differentially expressed in tumors.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF038451; AAC82614.1; -.
DR EMBL; AF007791; AAC77358.1; -.
DR EMBL; AF088867; AAF22484.1; -.
DR EMBL; BC015503; AAH15503.1; -.
SQ SEQUENCE 175 AA; 19979 MW; F271B1BD377BEE11 CRC64;

```
  Query Match           84.6%;  Score 899;  DB 4;  Length 175;
  Best Local Similarity  100.0%;  Pred. No. 4.1e-80;
  Matches  175;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      32 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 91
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 60

Qy      92 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 151
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 120

Qy     152 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 206
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 175


RESULT 2
O88312
ID   O88312      PRELIMINARY;      PRT;   175 AA.
AC   O88312;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   GOB-4 PROTEIN (ANTERIOR GRADIENT 2) (XENEPUS LAEVIS) (XENOPUS
DE   LAEVIS).
GN   AGR2 OR GOB-4 OR MAG-2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C; TISSUE=INTESTINE;
RA   Komiya T., Hirohashi S.;
RT   "Cloning of a gene, gob-4, which is expressed in intestinal goblet
RT   cells in mice.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SWISS WEBSTER;
RX   MEDLINE=99009231; PubMed=9790916;
RA   Thompson D.A., Weigel R.J.;
RT   "hAG-2, the human homologue of the Xenopus laevis cement gland gene
RT   XAG-2, is coexpressed with estrogen receptor in breast cancer cell
RT   lines.";
RL   Biochem. Biophys. Res. Commun. 251:111-116(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=PANCREAS;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=COLON, NORMAL. 5 MONTH OLD MALE MOUSE.;
RA   Strausberg R.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB016592; BAA32044.1; -.
DR   EMBL; AF044262; AAC72705.1; -.
DR   EMBL; AK007677; BAB25181.1; -.
DR   EMBL; BC013334; AAH13334.1; -.
DR   MGD; MGI:1344405; Agr2.
SQ   SEQUENCE   175 AA;  19920 MW;  ACC3CFE429B668CA CRC64;

  Query Match           77.9%;  Score 828;  DB 11;  Length 175;
  Best Local Similarity  91.4%;  Pred. No. 3.6e-73;
  Matches  160;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      32 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 91
           |||  ||| |||||:| |||:|||||  |||||  ||||||||||||||||||||||||||
Db       1 MEKFSVSAILLLVAISGTLAKDTTVKSGAKKDPKDSRPKLPQTLSRGWGDQLIWTQTYEE 60

Qy      92 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 151
           |||:|||||:|||:|||||||||||||||||||:|||||||||||||||||||||||||||
Db      61 ALYRSKTSNRPLMVIHHLDECPHSQALKKVFAEHKEIQKLAEQFVLLNLVYETTDKHLSP 120

Qy     152 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 206
           ||||||||:|||||||||||||||||||||||||:||||| ||||||||||||||
Db     121 DGQYVPRIVFVDPSLTVRADITGRYSNRLYAYEPSDTALLYDNMKKALKLLKTEL 175


RESULT 3
Q90Y05
ID   Q90Y05      PRELIMINARY;      PRT;   159 AA.
AC   Q90Y05;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CEMENT GLAND-SPECIFIC PROTEIN CGS.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Alexandrova E.M., Novoselov V.V., Zaraisky A.G.;
RT   "Three novel genes expressed in the anterior part of the Xenopus
RT   laevis embryo.";
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF314056; AAL26844.1; -.
SQ   SEQUENCE   159 AA;  18275 MW;  738D62284838B8EB CRC64;

  Query Match           51.6%;  Score 548;  DB 13;  Length 159;
  Best Local Similarity  60.0%;  Pred. No. 8e-46;
  Matches  105;  Conservative   26;  Mismatches   28;  Indels   16;  Gaps    1;

Qy      32 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 91
           ||  :  | | |||| |:|||::                : ||||||||| | | |||||
Db       1 METVLKSLFFLLVATSFTLAKE---------------RKPQTLSRGWGDNLEWVQTYEE 44

Qy      92 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 151
             |:|:|: ||||::|:|| ::|||||||| |||:: |||||:|:|||:||: |||:|
Db      45 GLFKAKSENKPLLLINHRNDCPHSQALKKAFAEHQGIQKLAEEFILLNVVYDPTDKNLQL 104

Qy     152 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 206
           ||||||:::|||||| |||| |:||| | |||||   | :|||||| ||||||
Db     105 DGQYVPKVVFVDPSLVVRADLPGKYSNHQYTYEPADIDHLFENMKKALVLLKTEL 159


RESULT 4
O42251
ID   O42251      PRELIMINARY;      PRT;   185 AA.
AC   O42251;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
```

```
DE   CEMENT GLAND-SPECIFIC.
GN   XAG-2.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=HEAD;
RX   MEDLINE=89324051; PubMed=2752418;
RA   Sive H.L., Hattori K., Weintraub H.;
RT   "Progressive determination during formation of the anteroposterior
RT   axis in Xenopus laevis.";
RL   Cell 58:171-180(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=HEAD;
RA   Sive H.L., Wainstock D.H., Kennedy B.S.;
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF025474; AAB81968.1; -.
SQ   SEQUENCE   185 AA;  20498 MW;  D6CDE02DEC3B857B CRC64;

  Query Match             42.1%;  Score 447;  DB 13;  Length 185;
  Best Local Similarity   50.6%;  Pred. No. 7.5e-36;
  Matches   89;  Conservative   26;  Mismatches   49;  Indels   12;  Gaps    2;

Qy      43 LVALSYTLARDTTVKP-----------GAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 91
           || |   |      ||            || :    : |  :|| |||| : | |||||
Db      10 LVLLCSVLGEAALRKPKRQAGATDTNGAAKSEPAPVKTKGLKTLDRGWGEDIEWAQTYEE 69

Qy      92 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLA-EQFVLLNLVYETTDKHLS 150
             | |:: :|||||:||||::||:| |||| |  :|  |||| | |::|||||:   |:: |
Db      70 GLAKARENNKPLMVIHHLEDCPYSIALKKAFVADKMAQKLAQEDFIMLNLVHPVADENQS 129

Qy     151 PDGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 206
           ||| ||||::|:|||||||:|: ||| |:|||:  |   |: |||||   |||||
Db     130 PDGHYVPRVIFIDPSLTVRSDLKGRYGNKLYAYDADDIPELITNMKKAKSFLKTEL 185


RESULT 5
Q96H50
ID   Q96H50      PRELIMINARY;      PRT;   172 AA.
AC   Q96H50;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 19.2 KDA PROTEIN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=OVARY, AND ADENOCARCINOMA;
RA   Strausberg R.;
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC008913; AAH08913.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   172 AA;  19228 MW;  5C9DDA6D5A7C4BDF CRC64;

  Query Match             21.4%;  Score 227;  DB 4;  Length 172;
  Best Local Similarity   34.8%;  Pred. No. 2.3e-14;
  Matches   55;  Conservative   25;  Mismatches   56;  Indels   22;  Gaps    3;

Qy      66 DSRPKLPQT------------------LSRGWGDQLIWTQTYEEALYKSKTSNKPLMII 106
           ::||:|  |                  | :|:|| : | :| |:   ::  |  |||:|
Db       2 ETRPRLGATCLLGFSFLLLVISSDGHNGLGKGFGDHIHW-RTLEDGKKEAAASGLPLMVI 60

Qy     107 HHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKH--LSPDGQYVPRIMFVDP 164
             |   |   :|||  |||: || :|:  ||::||  |  .||   |||| |:|||:|:||
Db      61 IHKSWCGACKALKPKFAESTEISELSHNFVMVNLEDEEEPKHEDFSPDGGYIPRILFLDP 120

Qy     165 SLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLL 202
           |  |  :|     |  | |       ::  ||:| : |
Db     121 SGKVHPEIINENGNPSYKYFYVSAEQVVQGMKEAQERL 158


RESULT 6
Q9CQU0
ID   Q9CQU0      PRELIMINARY;      PRT;   170 AA.
AC   Q9CQU0;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   0610040B21RIK PROTEIN (RIKEN CDNA 0610040B21 GENE).
GN   0610040B21RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=EMBRYO, AND KIDNEY;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK003481; BAB22811.1; -.
DR   EMBL; AK002862; BAB22413.1; -.
DR   EMBL; BC006857; AAH06857.1; -.
DR   MGD; MGI:1913323; 0610040B21Rik.
DR   InterPro; IPR000886; ER_target.
DR   InterPro; IPR000063; Thiored.
DR   PROSITE; PS00014; ER_TARGET; UNKNOWN_1.
SQ   SEQUENCE   170 AA;  19048 MW;  5B91FC9BE12C5E44 CRC64;

  Query Match             21.1%;  Score 224.5;  DB 11;  Length 170;
  Best Local Similarity   34.9%;  Pred. No. 4e-14;
  Matches   58;  Conservative   25;  Mismatches   58;  Indels   25;  Gaps    4;

Qy      39 AFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEEALYKSKT 98
           :||||:  |                  | |      | :|:|| : | :| |:   ::
Db      14 SFLLLITSS-----------------DGR----TGLGKGFGDHIHW-RTLEDGKKEAAA 50

Qy      99 SNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNL--VYETTDKHLSPDGQYV 156
           |  |||:| |   |   :|||  |||: || :|:  ||::||    |  |:  ||||  |:
Db      51 SGLPLMVIIHKSWCGACKALKPKFAESTEISELSHNFVMVNLEDEEEPRDEDFSPDGGYI 110

Qy     157 PRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLL 202
           |||:|:|||  || :|     |  | |       ::  ||:| : |
Db     111 PRILFLDPSGKVRPEIINESGNPSYKYFYVSAEQVVQGMKEAQERL 156


RESULT 7
```

```
O95881
ID   O95881      PRELIMINARY;      PRT;   172 AA.
AC   O95881;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   UNKNOWN (HYPOTHETICAL 19.2 KDA PROTEIN).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96207227; PubMed=8619474;
RA   Andersson B., Wentland M.A., Ricafrente J.Y., Liu W., Gibbs R.A.;
RT   "A 'double adaptor' method for improved shotgun library
RT   construction.";
RL   Anal. Biochem. 236:107-113(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97264341; PubMed=9110174;
RA   Yu W., Andersson B., Worley K.C., Muzny D.M., Ding Y., Liu W.,
RA   Ricafrente J.Y., Wentland M.A., Lennon G., Gibbs R.A.;
RT   "Large-scale concatenation cDNA sequencing.";
RL   Genome Res. 7:353-358(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Mei G., Yu W., Gibbs R.A.;
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=KIDNEY ADENOCARCINOMA;
RA   Strausberg R.;
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RC   TISSUE=COLON ADENOCARCINOMA;
RA   Strausberg R.;
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF131758; AAD20035.1; -.
DR   EMBL; BC001493; AAH01493.1; -.
DR   EMBL; BC008953; AAH08953.1; -.
DR   InterPro; IPR000063; Thiored.
DR   PROSITE; PS00194; THIOREDOXIN; UNKNOWN_1.
KW   Hypothetical protein.
SQ   SEQUENCE   172 AA;  19206 MW;  3092E9515A7C4094 CRC64;

  Query Match             20.9%;  Score 222;  DB 4;  Length 172;
  Best Local Similarity   34.2%;  Pred. No. 7.1e-14;
  Matches   54;  Conservative   26;  Mismatches   56;  Indels   22;  Gaps    3;

Qy      66 DSRPKLPQT------------------LSRGWGDQLIWTQTYEEALYKSKTSNKPLMII 106
           ::||:|  |                  | :|:|| : | :| |:   ::  |  |||:|
Db       2 ETRPRLGATCLLGFSFLLLVISSDGHNGLGKGFGDHIHW-RTLEDGKKEAAASGLPLMVI 60

Qy     107 HHLDECPHSQALKKVFAENKEIQKLAEQFVLLNL--VYETTDKHLSPDGQYVPRIMFVDP 164
             |   |   :|||  |||: || :|:  ||::||    |  |:  ||||  |:|||:|:||
Db      61 IHKSWCGACKALKPKFAESTEISELSHNFVMVNLEDEEEPKDEDFSPDGGYIPRILFLDP 120

Qy     165 SLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLL 202
           |  |  :|     |  | |       ::  ||:| : |
Db     121 SGKVHPEIINENGNPSYKYFYVSAEQVVQGMKEAQERL 158


RESULT 8
Q9NA78
ID   Q9NA78      PRELIMINARY;      PRT;   186 AA.
AC   Q9NA78;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Y57A10A.23 PROTEIN.
GN   Y57A10A.23.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Smye R.;
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; AL117195; CAB55026.1; -.
DR   InterPro; IPR000063; Thiored.
SQ   SEQUENCE   186 AA;  21467 MW;  F94E13D2F773321E CRC64;

  Query Match             18.7%;  Score 199;  DB 5;  Length 186;
  Best Local Similarity   30.0%;  Pred. No. 1.4e-11;
  Matches   51;  Conservative   38;  Mismatches   57;  Indels   24;  Gaps    6;

Qy      41 LLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEEALYKSKTSN 100
           |||:||         |   |       :  :   |:||:|| : | : :|:|:  :  ::
Db       4 LLLLAL---------VSASAYASFDKIKDSIQNPLARGFGDDIAWVK-WEDAIETALDTD 53

Qy     101 KPLMIIHHLDECPHSQALKKVFAEN---KEIQKLAEQFVLLNLV--YETTDKHLSPDGQY 155
           ||: :: |   |   :|||| | ::   |  :||:| ||::|     |  ::   |||:|
Db      54 KPIFLLIHKSWCHACKALKKTFQQSNAKKAFKKLSEHFVMVNTEDDDEPFEEEYRPDGKY 113

Qy     156 VPRIMFVDPSLTVRADITGRYSNRL-----YAYEPADTALLLDNMKKALK 200
           :||::|:| :     |:   : |:      |||  :  | :|::||  ||
Db     114 IPRLLFLDKN----GDLLQEFKNKKAEYKNYAYYYSSPADILNSMKDVLK 159


RESULT 9
Q9N5S7
ID   Q9N5S7      PRELIMINARY;      PRT;   257 AA.
AC   Q9N5S7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 30.2 KDA PROTEIN.
GN   F49H12.5.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Wohldmann P., Nhan M., Chissoe S., Hawkins J.;
RT   "The sequence of C. elegans cosmid F49H12.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RT   "Direct Submission.";
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC006642; AAF39830.1; -.
DR   InterPro; IPR000886; ER_target.
DR   InterPro; IPR000063; Thiored.
DR   PROSITE; PS00014; ER_TARGET; UNKNOWN_1.
```

```
KW   Hypothetical protein.
SQ   SEQUENCE   257 AA;  30205 MW;  BF7FBA62D04DBED6 CRC64;

  Query Match             15.5%;  Score 165;  DB 5;  Length 257;
  Best Local Similarity   30.7%;  Pred. No. 4.6e-08;
  Matches   51;  Conservative   35;  Mismatches   58;  Indels   22;  Gaps    8;

Qy      41 LLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEEALYKSKTSN 100
           |||::|   ||     |  :|:: ||        || |:   : | : :::|:  :|  |
Db       3 LLLLSL---LAVAVLAKKESKEEIKD--------LSHGYPSAIEWVE-FDKAIGIAKDLN 50

Qy     101 KPLMIIHHLDECPHSQALK---KVFAENKEIQKLAEQFVLLNLV--YETTDKHLSPDGQY 155
           ||:  : |   |   ::||   |   :  |:  |: :||::|:    |  |   |||| |
Db      51 KPIFFLIHKTWCGACKSLKRELKSSPKTDELIILSRKFVMVNVEDDEEPEDPKYSPDGGY 110

Qy     156 VPRIMFVD-PSLTVRADITGRYSNRLYAYE-PADTALLLDNMKKAL 199
           :|||:|:|    :: :   :| |  | |  ||    ::| |::||
Db     111 IPRILFLDTDGNPLKTNNERKYKNNKYFYPLPAQ---IIDGMERAL 153


RESULT 10
O27777
ID   O27777      PRELIMINARY;      PRT;   150 AA.
AC   O27777;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   PROTEIN DISULPHIDE ISOMERASE.
GN   MTH1745.
OS   Methanothermobacter thermautotrophicus.
OC   Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC   Methanothermobacter.
OX   NCBI_TaxID=145262;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DELTA H;
RX   MEDLINE=98037514; PubMed=9371463;
RA   Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., DuBois J.,
RA   Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA   Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA   Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA   Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA   McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA   Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;
RT   "Complete genome sequence of Methanobacterium thermoautotrophicum
RT   deltaH: functional analysis and comparative genomics.";
RL   J. Bacteriol. 179:7135-7155(1997).
DR   EMBL; AE000930; AAB86215.1; -.
DR   InterPro; IPR000063; Thiored.
DR   Pfam; PF00085; thiored; 1.
DR   PROSITE; PS00194; THIOREDOXIN; UNKNOWN_1.
KW   Isomerase; Complete proteome.
SQ   SEQUENCE   150 AA;  16939 MW;  E0672E64B7E340EB CRC64;

  Query Match             11.2%;  Score 119;  DB 17;  Length 150;
  Best Local Similarity   27.1%;  Pred. No. 0.00071;
  Matches   35;  Conservative   28;  Mismatches   46;  Indels   20;  Gaps    4;

Qy      39 AFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEEALYKSKT 98
           | | || : ||:                 :|::||:|:    | | |   ::||: ::
Db      19 AGLSLVLIIYTV----------------QPRVPQSLTTDEKD-LKWYTEHDEAIKEASR 60

Qy      99 SNKPLMIIHHLDECPHSQALKKVFAENKEIQ-KLAEQFVLLNLVYETTDKHLSPDGQY-V 156
           : | : ::     ||  | |:    :| |:| :||| |: : :  :|:    |    | |
Db      61 TGKNVFMVFSASWCPACQKLESETLQNTEVQRRLAEDFIAVKIDVDTSPALSSRYRIYGV 120

Qy     157 PRIMFVDPS 165
           | :: :|||
Db     121 PTVIILDPS 129


RESULT 11
Q924S5
ID   Q924S5      PRELIMINARY;      PRT;   950 AA.
AC   Q924S5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   LON.
GN   LON.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Hori O., Ogawa S.;
RT   "rat Lon.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB064323; BAB62423.1; -.
SQ   SEQUENCE   950 AA;  105792 MW;  D505C3D851B6F0E7 CRC64;

  Query Match             8.4%;  Score 89;  DB 11;  Length 950;
  Best Local Similarity   27.5%;  Pred. No. 7.2;
  Matches   46;  Conservative   27;  Mismatches   72;  Indels   22;  Gaps    8;

Qy      50 LARDTTVKPGAKKDTKDSRPKLPQTLSRG---WGDQLIWTQTYE---EALYKSKTSNKPL 103
           ::|   |:|   :   ::: |  : | ||    ||:|      |   ||   | || : |
Db     204 ISRQLEVEPEGLEPEAENKQKSRRKLKRGKKEVGDELGAKPQLEMVTEA--TSDTSKEVL 261

Qy     104 MIIHHLDECPHS--QALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSPDGQYVPRIMF 161
           |:   ::    |   |  ::| |   || |     : || :|  :   :   || |
Db     262 MV--EVENVAHEDFQVTEEVKALTAEIVKTIRDIIALNPLYRESVLQMMQAGQRV----- 314

Qy     162 VDPSL---TVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTE 205
           ||  :    : | :||  |: |   :  :   :|  : ||| ||| |
Db     315 VDNPIYLSDMGAALTGAESHEL--QDVLEETNILKRLYKALSLLKKE 359


RESULT 12
Q9AF04
ID   Q9AF04      PRELIMINARY;      PRT;   737 AA.
AC   Q9AF04;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   HYPOTHETICAL 78.0 KDA PROTEIN.
OS   Frankia sp. CpI1.
OG   Plasmid pFQ12.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Frankineae; Frankiaceae; Frankia.
OX   NCBI_TaxID=1856;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   John T.R., Rice J.M., Johnson J.D.;
RT   "Analysis of pFQ12; a 22.4 kb Frankia plasmid.";
RL   Can. J. Microbiol. 0:0-0(2001).
DR   EMBL; AY027524; AAK20150.1; -.
DR   InterPro; IPR002543; FtsK_SpoIIIE.
DR   Pfam; PF01580; FtsK_SpoIIIE; 1.
KW   Hypothetical protein; Plasmid.
SQ   SEQUENCE   737 AA;  78018 MW;  DEFCFB45DFDB7C0F CRC64;

  Query Match             8.3%;  Score 88.5;  DB 2;  Length 737;
  Best Local Similarity   26.3%;  Pred. No. 5.8;
  Matches   46;  Conservative   24;  Mismatches   70;  Indels   35;  Gaps    8;

Qy      43 LVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSR----------GWGDQLIWTQTYEEA 92
           |   |  || |  :  ||  :        || |  |          |: : |::  | |
Db     201 LARFSEDLANDADLPAGAGVEV------LPGTTRRSAVLEISTVDGFAEDLLFGDDYSEL 254

Qy      93 LYKSKTSNKPLMIIHHLDECPHSQALKKVFA---ENKEIQKLAEQFVLLNLVYETTDK-- 147
                | | ||||  : :  |    |:: :|     :   |  : :||   :   ||
Db     255 -----TINNPLMIGRYRNTQPAEAHLRETYALVVGEQGSGKTNQLYVLTGQLARCTDVIV 309

Qy     148 -HLSPDGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPAD----TALLLDNMKK 197
```

```
          |: |:|  :||  :| | :  :|    | :    ||||:    |  ||  | :
Db    310 CHIDPNGGGLPR-PWVMPWVEGQAT---RPAVDWVAYEPVEAERMTGALLRGMTR 360


RESULT 13
Q9Z6X3
ID   Q9Z6X3      PRELIMINARY;      PRT;   348 AA.
AC   Q9Z6X3;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PREDICTED DISULFIDE BOND ISOMERASE.
GN   CPN0933 OR CPJ0933 OR CP0928.
OS   Chlamydia pneumoniae (Chlamydophila pneumoniae).
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.
OX   NCBI_TaxID=83558;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CWL029;
RX   MEDLINE=99206606; PubMed=10192388;
RA   Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA   Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT   "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL   Nat. Genet. 21:385-389(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AR39;
RX   MEDLINE=20150255; PubMed=10684935;
RA   Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA   White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA   Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA   Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA   Eisen J., Fraser C.M.;
RT   "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT   pneumoniae AR39.";
RL   Nucleic Acids Res. 28:1397-1406(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=J138;
RX   MEDLINE=20330349; PubMed=10871362;
RA   Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA   Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT   "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT   from Japan and CWL029 from USA.";
RL   Nucleic Acids Res. 28:2311-2314(2000).
DR   EMBL; AE001673; AAD19071.1; -.
DR   EMBL; AE002251; AAF38711.1; -.
DR   EMBL; AP002548; BAA99141.1; -.
DR   PHCI-2DPAGE; Q9Z6X3; -.
DR   TIGR; CP0928; -.
DR   InterPro; IPR000063; Thiored.
KW   Isomerase; Complete proteome.
SQ   SEQUENCE   348 AA;  40492 MW;  A7E03C043D775CEC CRC64;

  Query Match             8.2%;  Score 87;  DB 16;  Length 348;
  Best Local Similarity   25.0%;  Pred. No. 3;
  Matches   29;  Conservative   26;  Mismatches   37;  Indels   24;  Gaps    4;

Qy      29 RVAMEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQT 88
           | ::::: | | ||   || :|    |::    :|:               | ::|
Db       9 RCSLKQLKVLATLL---LSLSL---PTLEAAENRDS---------------DSIVWHLD 46

Qy      89 YEEALYKSKTSNKPLMIIHHLDEC--PHSQALKKVFAENKEIQKLAEQFVLLNLVY 142
           |:||| ||| :  ||::|    :   |  :  |:|    : |::: :|| : : |
Db      47 YQEALQKSKEAELPLLVIFSGSDWNGPCMKIRKEVLESPEFIKRVQGKFVCVEVEY 102


RESULT 14
Q929J2
ID   Q929J2      PRELIMINARY;      PRT;  1151 AA.
AC   Q929J2;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PUTATIVE PEPTIDOGLYCAN BOUND PROTEIN (LPXTG MOTIF).
GN   LIN2283.
OS   Listeria innocua.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Listeria.
OX   NCBI_TaxID=1642;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CLIP 11262 / SEROVAR 6A;
RX   PubMed=11679669;
RA   Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA   Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA   Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA   Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA   Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA   Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA   Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA   Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA   Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA   Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA   Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT   "Comparative genomics of Listeria species.";
RL   Science 294:849-852(2001).
DR   EMBL; AL596171; CAC97511.1; -.
DR   ListiList; LIN02283; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   1151 AA;  127525 MW;  6A65870E5FBF6E1E CRC64;

  Query Match             7.9%;  Score 84.5;  DB 16;  Length 1151;
  Best Local Similarity   22.8%;  Pred. No. 26;
  Matches   44;  Conservative   31;  Mismatches   51;  Indels   67;  Gaps   10;

Qy      50 LARDTTVKPGAKK--DTKDSRPKLPQTLSRGWGDQLIWTQ----------------TYE 90
           :| | :| |   :  |    :   |:||:  :|::    |                 |:
Db     961 VALDDSVAPVTVQYVDQNHKQIASPETLTGAYGEKFTAKQKKITNYTLVKTPANVSGTFN 1020

Qy      91 EA------LYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVL---LNLV 141
           |       :|: '|:    :|: ::|               |  :|||:  ||   ||
Db    1021 EKAQTVTFVYQKVTAGN--IIVDYVD--------------KNGEKLADSIVLTGKLNSS 1063

Qy     142 YETTDKHLSPDGQY-VPRIMFVDPSLTVRADITGRYSNR----LYAYEPADTALL----- 191
           | |: | :|    |  |:            : :|::||      | || | |  :
Db    1064 YRTSAKKISGYKLYQTPK-----------NASGKFSNTSQRVTYVYEKASTFSISSNKG 1111

Qy     192 LDNMKKALKLLKT 204
           ||::|: || :|
Db    1112 ADNVEKSAKLPQT 1124


RESULT 15
Q9BJK8
ID   Q9BJK8      PRELIMINARY;      PRT;   435 AA.
AC   Q9BJK8;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   NUCLEAR RECEPTOR NHR-79 (FRAGMENT).
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Bogan A., Maina C.V., Yamamoto K., Cohen F., Sluder A.E.;
RT   "Caenorhabditis elegans nuclear receptor sequences exhibit biophysical
RT   compatability with the ligand-binding domain fold.";
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR   EMBL; AF332208; AAK17979.1; -.
DR   HSSP; P20393; 1A6Y.
DR   InterPro; IPR000536; Hormone_rec_lig.
```

```
DR   InterPro; IPR001628; zf-C4.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
KW   DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW   Zinc-finger.
FT   NON_TER       1      1
SQ   SEQUENCE   435 AA;  50536 MW;  B946B47CE5AE1492 CRC64;

  Query Match             7.9%;  Score 84;  DB 5;  Length 435;
  Best Local Similarity   22.1%;  Pred. No. 7.8;
  Matches   25;  Conservative   26;  Mismatches   50;  Indels   12;  Gaps    3;

Qy      89 YEEALYKSKTSNKPLMIIHHLDECPHSQALKK-----VFAENKEIQKLAEQFVLLNLVYE 143
           ::   |: |  |   ||: :::  |  ::| |     :|     :  | :|     :::
Db     215 FDHRTYRQKNRNDYTMILDYINRFPEFESLSKSEKTVLFRTAAAVDVLLDQSYYSQVIFP 274

Qy     144 TTDKHLSPDGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMK 196
           | |  :: :|:|:|    ::|   |      | |   :: |  |   :| :||
Db     275 TEDVLVTANGEYLP----MNPMPKVE---NQRDSGNFHSDEDYDRFKMLTSMK 320


Search completed: October  9, 2002, 16:41:11
Job time : 29 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 16:39:22 ; Search time 18 Seconds
(without alignments)
279.537 Million cell updates/sec

Title: US-09-674-266A-181
Perfect score: 1063
Sequence: 1 RLSCAGTLSGSGPHPSRRLT.....DTALLLDNMKKALKLLKTEL 206

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	899	84.6	175	4	US-08-916-576B-2	Sequence 2, Appli
2	899	84.6	175	4	US-09-247-155-106	Sequence 106, App
3	609.5	57.3	170	4	US-08-916-576B-8	Sequence 8, Appli
4	605	56.9	131	4	US-09-247-155-174	Sequence 174, App
5	555	52.2	166	4	US-08-916-576B-6	Sequence 6, Appli
6	448	42.1	183	4	US-08-916-576B-7	Sequence 7, Appli
7	222	20.9	172	4	US-08-916-576B-4	Sequence 4, Appli
8	81.5	7.7	901	2	US-08-884-681-5	Sequence 5, Appli
9	81.5	7.7	901	4	US-09-258-643-5	Sequence 5, Appli
10	81.5	7.7	968	3	US-08-560-005-7	Sequence 7, Appli
11	81.5	7.7	968	4	US-09-418-540-7	Sequence 7, Appli
12	78	7.3	129	4	US-09-534-638-5	Sequence 5, Appli
13	78	7.3	453	4	US-09-013-881-5	Sequence 5, Appli
14	78	7.3	1160	5	PCT-US92-05401-4	Sequence 4, Appli
15	77	7.2	652	2	US-08-313-185-53	Sequence 53, Appl
16	77	7.2	652	2	US-08-459-499-17	Sequence 17, Appl
17	77	7.2	652	3	US-09-082-614A-53	Sequence 53, Appl
18	74	7.0	264	3	US-09-053-197A-8	Sequence 8, Appli
19	74	7.0	264	4	US-09-085-761A-8	Sequence 8, Appli
20	71.5	6.7	352	3	US-08-921-887-52	Sequence 52, Appl
21	71	6.7	529	3	US-08-821-984-6	Sequence 6, Appli
22	71	6.7	529	3	US-08-821-984-8	Sequence 8, Appli
23	71	6.7	529	4	US-09-329-749-6	Sequence 6, Appli
24	71	6.7	529	4	US-09-329-749-8	Sequence 8, Appli
25	70.5	6.6	481	1	US-08-186-811-2	Sequence 2, Appli
26	70.5	6.6	481	1	US-08-311-611A-98	Sequence 98, Appl
27	70.5	6.6	481	1	US-08-372-783-98	Sequence 98, Appl
28	70.5	6.6	481	1	US-08-372-105-98	Sequence 98, Appl
29	70.5	6.6	481	1	US-08-306-473A-98	Sequence 98, Appl
30	70.5	6.6	481	1	US-08-261-660A-4	Sequence 4, Appli
31	70.5	6.6	481	1	US-08-209-762-98	Sequence 98, Appl
32	70.5	6.6	481	1	US-08-473-344-98	Sequence 98, Appl
33	70.5	6.6	481	1	US-08-274-303-6	Sequence 6, Appli
34	70.5	6.6	481	1	US-08-377-391A-2	Sequence 2, Appli
35	70.5	6.6	481	2	US-08-485-445A-98	Sequence 98, Appl
36	70.5	6.6	481	2	US-08-779-400-2	Sequence 2, Appli
37	70.5	6.6	481	2	US-08-955-660-2	Sequence 2, Appli
38	70.5	6.6	481	3	US-09-119-263-98	Sequence 98, Appl
39	70.5	6.6	481	4	US-08-657-162-98	Sequence 98, Appl
40	70.5	6.6	481	4	US-09-224-480-98	Sequence 98, Appl
41	70.5	6.6	481	4	US-09-093-539-98	Sequence 98, Appl
42	70.5	6.6	481	4	US-09-146-620-2	Sequence 2, Appli
43	70.5	6.6	481	4	US-09-395-453-2	Sequence 2, Appli
44	70.5	6.6	481	5	PCT-US94-02465-98	Sequence 98, Appl
45	70.5	6.6	481	5	PCT-US94-06931-4	Sequence 4, Appli

ALIGNMENTS

```
RESULT 1
US-08-916-576B-2
; Sequence 2, Application US/08916576B
; Patent No. 6171816
;  GENERAL INFORMATION:
;    APPLICANT:  YU, GUO-LIANG
;    APPLICANT:  DILLON, PATRICK J.
;    APPLICANT:  EBNER, REINHARD
;    APPLICANT:  ENDRESS, GREGORY A.
;    TITLE OF INVENTION:  NOVEL HUMAN GROWTH FACTORS
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  US
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/916,576B
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/024,347
;      FILING DATE:  23-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  STEFFE, ERIC K.
;      REGISTRATION NUMBER:  36,688
;      REFERENCE/DOCKET NUMBER:  1488.0500001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  175 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-916-576B-2

  Query Match             84.6%;  Score 899;  DB 4;  Length 175;
  Best Local Similarity   100.0%;  Pred. No. 8.5e-96;
  Matches  175;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      32 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 91
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 60

Qy     92 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 151
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 120

Qy    152 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 206
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 175


RESULT 2
US-09-247-155-106
; Sequence 106, Application US/09247155A*
; Patent No. 6312922
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, Jean-Baptiste
;  APPLICANT: Duclert, Aymeric
;  APPLICANT: Bougueleret, Lydie
;  TITLE OF INVENTION: Complementary DNAs
;  FILE REFERENCE: GENSET.021A
;  CURRENT APPLICATION NUMBER: US/09/247,155A
;  CURRENT FILING DATE: 1999-02-09
;  EARLIER APPLICATION NUMBER: 60/074,121
;  EARLIER FILING DATE: 1998-02-09
;  EARLIER APPLICATION NUMBER: 60/081,563
;  EARLIER FILING DATE: 1998-04-13
;  EARLIER APPLICATION NUMBER: 60/096,116
;  EARLIER FILING DATE: 1998-08-10
;  EARLIER APPLICATION NUMBER: 60/099,273
;  EARLIER FILING DATE: 1998-10-04
;  NUMBER OF SEQ ID NOS: 182
;  SOFTWARE: Patent.pm
; SEQ ID NO 106
;   LENGTH: 175
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SIGNAL
;   LOCATION: -20..-1
US-09-247-155-106

  Query Match             84.6%;  Score 899;  DB 4;  Length 175;
  Best Local Similarity   100.0%;  Pred. No. 8.5e-96;
  Matches  175;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     32 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 91
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 60

Qy     92 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 151
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 120

Qy    152 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 206
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 175


RESULT 3
US-08-916-576B-8
; Sequence 8, Application US/08916576B
; Patent No. 6171816
;  GENERAL INFORMATION:
;    APPLICANT:  YU, GUO-LIANG
;    APPLICANT:  DILLON, PATRICK J.
;    APPLICANT:  EBNER, REINHARD
;    APPLICANT:  ENDRESS, GREGORY A.
;    TITLE OF INVENTION:  NOVEL HUMAN GROWTH FACTORS
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  US
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/916,576B
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/024,347
;      FILING DATE:  23-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  STEFFE, ERIC K.
;      REGISTRATION NUMBER:  36,688
;      REFERENCE/DOCKET NUMBER:  1488.0500001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  170 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-916-576B-8

  Query Match             57.3%;  Score 609.5;  DB 4;  Length 170;
  Best Local Similarity   71.3%;  Pred. No. 1.9e-62;
  Matches  122;  Conservative   14;  Mismatches   26;  Indels    9;  Gaps    2;

Qy     37 VSAFLLLVALSYTLARDTTVKPGAKKDTKDSRP-KLPQTLSRGWGDQLIWTQTYEEALYK 95
          : | |||||||  ||                 || || | ||||||||||| : | ||||| | |
Db      7 LGACLLLVALSXXLA--------XXXKGKDXRPIKGPQTLSRGWGDXIXWVQTYEEGLXK 58

Qy     96 SKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSPDGQY 155
          :  |||||:||||::||:||||||||||| |||:||: ||:||||:||||::|||||||
Db     59 AXXSNKPLMVIHHLEDCPYSQALKKVFAENXEIQELAQNFVMLNLVHETTDENLSPDGQY 118

Qy    156 VPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 206
          |||||||||||||||| ||| ||||||||  |  ||: ||||||||||||
Db    119 VPRIMFVDPSLTVRADIXGRYGNRLYAYEPXDXPLLIXNMKKALKLLKTEL 169


RESULT 4
US-09-247-155-174
; Sequence 174, Application US/09247155A
; Patent No. 6312922
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, Jean-Baptiste
;  APPLICANT: Duclert, Aymeric
;  APPLICANT: Bougueleret, Lydie
;  TITLE OF INVENTION: Complementary DNAs
;  FILE REFERENCE: GENSET.021A
;  CURRENT APPLICATION NUMBER: US/09/247,155A
;  CURRENT FILING DATE: 1999-02-09
;  EARLIER APPLICATION NUMBER: 60/074,121
;  EARLIER FILING DATE: 1998-02-09
;  EARLIER APPLICATION NUMBER: 60/081,563
;  EARLIER FILING DATE: 1998-04-13
;  EARLIER APPLICATION NUMBER: 60/096,116
;  EARLIER FILING DATE: 1998-08-10
;  EARLIER APPLICATION NUMBER: 60/099,273
;  EARLIER FILING DATE: 1998-10-04
;  NUMBER OF SEQ ID NOS: 182
;  SOFTWARE: Patent.pm
; SEQ ID NO 174
```

```
;   LENGTH: 131
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SIGNAL
;   LOCATION: -20..-1
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: 40,41,43,60,70,76,82,86,105,107
;   OTHER INFORMATION: Xaa = any one of the twenty amino acids
US-09-247-155-174

  Query Match             56.9%;  Score 605;  DB 4;  Length 131;
  Best Local Similarity   92.9%;  Pred. No. 4.2e-62;
  Matches  117;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      32 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 91
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEX 60

Qy      92 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 151
           | ||||||||||||||||| ||||||||| ||||| ||||| ||| ||||||||||||||
Db      61 XLXKSKTSNKPLMIIHHLDXCPHSQALKKXFAENKXIQKLAXQFVXLNLVYETTDKHLSP 120

Qy     152 DGQYVP 157
           |||| |
Db     121 DGQYXP 126


RESULT 5
US-08-916-576B-6
; Sequence 6, Application US/08916576B
; Patent No. 6171816
;  GENERAL INFORMATION:
;    APPLICANT:  YU, GUO-LIANG
;    APPLICANT:  DILLON, PATRICK J.
;    APPLICANT:  EBNER, REINHARD
;    APPLICANT:  ENDRESS, GREGORY A.
;    TITLE OF INVENTION:  NOVEL HUMAN GROWTH FACTORS
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  US
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/916,576B
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/024,347
;      FILING DATE:  23-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  STEFFE, ERIC K.
;      REGISTRATION NUMBER:  36,688
;      REFERENCE/DOCKET NUMBER:  1488.0500001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  166 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-916-576B-6

  Query Match             52.2%;  Score 555;  DB 4;  Length 166;
  Best Local Similarity   64.7%;  Pred. No. 3.5e-56;
  Matches  108;  Conservative   28;  Mismatches   19;  Indels   12;  Gaps    4;

Qy      41 LLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEEALYKSKTSN 100
           |||| :|  ||    :|       |: ||  |||||||||| : | ||||| |: :: |
Db      11 LLLVTVSSNLA--IAIK-------KEKRP--PQTLSRGWGDDITWVQTYEEGLFYAQKSK 59

Qy     101 KPLMIIHHLDECPHSQALKKVFAENKEIQKLAE-QFVLLNLVYETTDKHLSPDGQYVPRI 159
           ||||:||||::| :|||||||||:|:|||::|: :|::|||::|||||:||||||||||
Db      60 KPLMVIHHLEDCQYSQALKKVFAQNEEIQEMAQNKFIMLNLMHETTDKNLSPDGQYVPRI 119

Qy     160 MFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 206
           |||||||||||| |||||||| ||| |  ||::||||||:|:::||
Db     120 MFVDPSLTVRADIAGRYSNRLYTYEPRDLPLLIENMKKALRLIQSEL 166


RESULT 6
US-08-916-576B-7
; Sequence 7, Application US/08916576B
; Patent No. 6171816
;  GENERAL INFORMATION:
;    APPLICANT:  YU, GUO-LIANG
;    APPLICANT:  DILLON, PATRICK J.
;    APPLICANT:  EBNER, REINHARD
;    APPLICANT:  ENDRESS, GREGORY A.
;    TITLE OF INVENTION:  NOVEL HUMAN GROWTH FACTORS
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  US
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/916,576B
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/024,347
;      FILING DATE:  23-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  STEFFE, ERIC K.
;      REGISTRATION NUMBER:  36,688
;      REFERENCE/DOCKET NUMBER:  1488.0500001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  183 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-916-576B-7

  Query Match             42.1%;  Score 448;  DB 4;  Length 183;
  Best Local Similarity   49.4%;  Pred. No. 8.6e-44;
  Matches   88;  Conservative   29;  Mismatches   51;  Indels   10;  Gaps    3;

Qy      39 AFLLLVALSYTLARDTTVKP---GAKKDTKDSRPKLP------QTLSRGWGDQLIWTQTY 89
           : : || |   |       ||       |||  :   |      :|| ||||: : | |||
Db       6 SLVCLVLLCSALGEAVLKKPKKQAGTTDTKTDQEPAPIKTKGLKTLDRGWGESIEWVQTY 65

Qy      90 EEALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLA-EQFVLLNLVYETTDKH 148
```

```
          || | |:: :|||||:||||::||:| |||| |  ::  |||| | |::||||:   |::
Db     66 EEGLAKARENNKPLMVIHHLEDCPYSIALKKAFVADRMAQKLAQEDFIMLNLVHPVADEN 125

Qy    149 LSPDGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 206
          |||| ||||::|:|||||||:|: ||| |::|||:  |   |: |||||   |||||
Db    126 QSPDGHYVPRVIFIDPSLTVRSDLKGRYGNKMYAYDADDIPELITNMKKAKSFLKTEL 183


RESULT 7
US-08-916-576B-4
; Sequence 4, Application US/08916576B
; Patent No. 6171816
;  GENERAL INFORMATION:
;    APPLICANT:  YU, GUO-LIANG
;    APPLICANT:  DILLON, PATRICK J.
;    APPLICANT:  EBNER, REINHARD
;    APPLICANT:  ENDRESS, GREGORY A.
;    TITLE OF INVENTION:  NOVEL HUMAN GROWTH FACTORS
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  US
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/916,576B
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/024,347
;      FILING DATE:  23-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  STEFFE, ERIC K.
;      REGISTRATION NUMBER:  36,688
;      REFERENCE/DOCKET NUMBER:  1488.0500001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  172 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-916-576B-4

  Query Match             20.9%;  Score 222;  DB 4;  Length 172;
  Best Local Similarity   34.2%;  Pred. No. 8.8e-18;
  Matches   54;  Conservative   26;  Mismatches   56;  Indels   22;  Gaps    3;

Qy     66 DSRPKLPQT------------------LSRGWGDQLIWTQTYEEALYKSKTSNKPLMII 106
          ::||:|  |                  | :|:|| : | :| |:   ::  |  |||:|
Db      2 ETRPRLGATCLLGFSFLLLVISSDGHNGLGKGFGDHIHW-RTLEDGKKEAAASGLPLMVI 60

Qy    107 HHLDECPHSQALKKVFAENKEIQKLAEQFVLLNL--VYETTDKHLSPDGQYVPRIMFVDP 164
          |   |   :|||  |||: || :|:  ||::||    |  |:  |||| |:|||:|:||
Db     61 IHKSWCGACKALKPKFAESTEISELSHNFVMVNLEDEEEPKDEDFSPDGGYIPRILFLDP 120

Qy    165 SLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLL 202
          |  |  :|     |  | |       ::  ||:| : |
Db    121 SGKVHPEIINENGNPSYKYFYVSAEQVVQGMKEAQERL 158


RESULT 8
US-08-884-681-5
; Sequence 5, Application US/08884681
; Patent No. 5955338
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Corley, Neil C.
;    APPLICANT:  Shah, Purvi
;    TITLE OF INVENTION:  NEW HUMAN PHOSPHATIDYLINOSITOL
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/884,681
;      FILING DATE:  Filed Herewith
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0334 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  901 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  GenBank
;      CLONE:  1420920
US-08-884-681-5

  Query Match             7.7%;  Score 81.5;  DB 2;  Length 901;
  Best Local Similarity   23.7%;  Pred. No. 1.6;
  Matches   45;  Conservative   32;  Mismatches   46;  Indels   67;  Gaps   13;

Qy     55 TVKPGAKKDTKDSRPK--LPQTLSRGWGDQLIWTQTYEEAL-YKS----KTS-NKPLMII 106
          | |  :| |  ||  |  :|      | |:::|  |    | |:|    ||| :||:  :
Db    476 TYKYDSKTDRWDSSGKCRVP-----AWCDRILWRGTNVNQLNYRSHMELKTSDHKPVSAL 530

Qy    107 HHL-----DECPHSQALKKVFAENKEIQKLAE-----------QFVLLNLVYETTDK-- 147
           |:     ||    :  :|||  ::  |    |           :||  |: :    |
Db    531 FHIGVKVVDE----RRYRKVFEDSVRIMDRMENDFLPSLELSRREFVFENVKFRQLQKGK 586

Qy    148 -HLSPDGQ------YVPRIM---------------FVDPSLTVRADITGRYSNRLYAYE 184
           :| :||      ::|::                 :::|: ||  ||:      |  |
Db    587 FQISNNGQVPCHFSFIPKLNDSQYCKPWLRAEPFEGYLEPNETV--DIS------LDVYV 638

Qy    185 PADTALLLDN 194
           |:  :|::
Db    639 SKDSVTILNS 648


RESULT 9
US-09-258-643-5
; Sequence 5, Application US/09258643
; Patent No. 6277373
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Corley, Neil C.
```

```
;    APPLICANT:  Shah, Purvi
;    TITLE OF INVENTION:  NEW HUMAN PHOSPHATIDYLINOSITOL
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/258,643
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/884,681
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0334 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  901 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  GenBank
;      CLONE:  1420920
US-09-258-643-5

  Query Match             7.7%;  Score 81.5;  DB 4;  Length 901;
  Best Local Similarity   23.7%;  Pred. No. 1.6;
  Matches   45;  Conservative   32;  Mismatches   46;  Indels   67;  Gaps   13;

Qy      55 TVKPGAKKDTKDSRPK--LPQTLSRGWGDQLIWTQTYEEAL-YKS----KTS-NKPLMII 106
           | |  :| |  ||  |  :|      | |:::|  |    | |:|    ||| :||:  :
Db     476 TYKYDSKTDRWDSSGKCRVP-----AWCDRILWRGTNVNQLNYRSHMELKTSDHKPVSAL 530

Qy     107 HHL-----DECPHSQALKKVFAENKEIQKLAE-----------QFVLLNLVYETTDK-- 147
            |:     ||    :  :|||  ::  |    |           :||  |: :    |
Db     531 FHIGVKVVDE----RRYRKVFEDSVRIMDRMENDFLPSLELSRREFVFENVKFRQLQKGK 586

Qy     148 -HLSPDGQ------YVPRIM---------------FVDPSLTVRADITGRYSNRLYAYE 184
            :| :||      ::|::                 :::|: ||  ||:      |  |
Db     587 FQISNNGQVPCHFSFIPKLNDSQYCKPWLRAEPFEGYLEPNETV--DIS------LDVYV 638

Qy     185 PADTALLLDN 194
             |:  :|::
Db     639 SKDSVTILNS 648


RESULT 10
US-08-560-005-7
; Sequence 7, Application US/08560005
; Patent No. 6001354
;  GENERAL INFORMATION:
;    APPLICANT:  Pot, David A.
;    APPLICANT:  Williams, Lewis T.
;    APPLICANT:  Jefferson, Anne Bennett
;    APPLICANT:  Majerus, Philip W.
;    TITLE OF INVENTION:  No. 6001354el Grb2 Associating Protein and Nucleic
;    TITLE OF INVENTION:  Acids Encoding Therefor
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew
;      STREET:  One Market Plaza, Steuart Tower, Suite 2000
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/560,005
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dow, Karen B.
;      REGISTRATION NUMBER:  29,684
;      REFERENCE/DOCKET NUMBER:  2307K-0624000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-326-2400
;      TELEFAX:  415-326-2422
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  968 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FEATURE:
;      NAME/KEY:  Region
;      LOCATION:  1..968
;      OTHER INFORMATION:  /note= "ocrl"
US-08-560-005-7

  Query Match             7.7%;  Score 81.5;  DB 3;  Length 968;
  Best Local Similarity   23.7%;  Pred. No. 1.8;
  Matches   45;  Conservative   32;  Mismatches   46;  Indels   67;  Gaps   13;

Qy      55 TVKPGAKKDTKDSRPK--LPQTLSRGWGDQLIWTQTYEEAL-YKS----KTS-NKPLMII 106
           | |  :| |  ||  |  :|      | |:::|  |    | |:|    ||| :||:  :
Db     551 TYKYDSKTDRWDSSGKCRVP-----AWCDRILWRGTNVNQLNYRSHMLLKTSDHKPVSAL 605

Qy     107 HHL-----DECPHSQALKKVFAENKEIQKLAE-----------QFVLLNLVYETTDK-- 147
            |:     ||    :  :|||  ::  |    |           :||  |: :    |
Db     606 FHIGVKVVDE----RRYRKVFEDSVRIMDRMENDFLPSLELSRREFVFENVKFRQLQKGK 661

Qy     148 -HLSPDGQ------YVPRIM---------------FVDPSLTVRADITGRYSNRLYAYE 184
            :| :||      ::|::                 :::|: ||  ||:      |  |
Db     662 FQISNNGQVPCHFSFIPKLNDSQYCKPWLRAEPFEGYLEPNETV--DIS------LDVYV 713

Qy     185 PADTALLLDN 194
             |:  :|::
Db     714 SKDSVTILNS 723


RESULT 11
US-09-418-540-7
; Sequence 7, Application US/09418540
; Patent No. 6296848
;  GENERAL INFORMATION:
;    APPLICANT:  Pot, David A.
;    APPLICANT:  Williams, Lewis T.
;    APPLICANT:  Jefferson, Anne Bennett
;    APPLICANT:  Majerus, Philip W.
;    TITLE OF INVENTION:  No. 6296848el Grb2 Associating Protein and Nucleic
;    TITLE OF INVENTION:  Acids Encoding Therefor
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew
;      STREET:  One Market Plaza, Steuart Tower, Suite 2000
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
```

```
;      ZIP:  94105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/418,540
;      FILING DATE:  14-OCT-1999
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/560,005
;      FILING DATE:  17-NOV-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dow, Karen B.
;      REGISTRATION NUMBER:  29,684
;      REFERENCE/DOCKET NUMBER:  2307K-0624000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-326-2400
;      TELEFAX:  415-326-2422
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  968 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FEATURE:
;      NAME/KEY:  Region
;      LOCATION:  1..968
;      OTHER INFORMATION:  /note= "ocr1"
US-09-418-540-7

  Query Match             7.7%;  Score 81.5;  DB 4;  Length 968;
  Best Local Similarity   23.7%;  Pred. No. 1.8;
  Matches   45;  Conservative   32;  Mismatches   46;  Indels   67;  Gaps   13;

Qy      55 TVKPGAKKDTKDSRPK--LPQTLSRGWGDQLIWTQTYEEAL-YKS----KTS-NKPLMII 106
           | |  :| |  ||  |  :|      | |:::|  |    | |:|    ||| :||:  :
Db     551 TYKYDSKTDRWDSSGKCRVP-----AWCDRILWRGTNVNQLNYRSHMLLKTSDHKPVSAL 605

Qy     107 HHL-----DECPHSQALKKVFAENKEIQKLAE-----------QFVLLNLVYETTDK-- 147
            |:     ||    :  :|:| ::  |    |            :||  |: :    |
Db     606 FHIGVKVVDE----RRYRKVFEDSVRIMDRMENDFLPSLELSRREFVFENVKFRQLQKGK 661

Qy     148 -HLSPDGQ------YVPRIM---------------FVDPSLTVRADITGRYSNRLYAYE 184
            :| :||      ::|::                :::|: ||  ||:      |  |
Db     662 FQISNNGQVPCHFSFIPKLNDSQYCKPWLRAEPFEGYLEPNETV--DIS------LDVYV 713

Qy     185 PADTALLLDN 194
           |:  :|::
Db     714 SKDSVTILNS 723


RESULT 12
US-09-534-638-5
; Sequence 5, Application US/09534638
; Patent No. 6320038
; GENERAL INFORMATION:
;  APPLICANT: Panula, Pertti A.J.
;  APPLICANT: Brandt, Annika
;  APPLICANT: Westerlund, Johanna
;  TITLE OF INVENTION: Promoter for Neuropeptide FF Promoter and use thereof
;  TITLE OF INVENTION: for therapy and diagnosis
;  FILE REFERENCE: 2530-104
;  CURRENT APPLICATION NUMBER: US/09/534,638
;  CURRENT FILING DATE: 2000-03-27
;  EARLIER APPLICATION NUMBER: 09/365755
;  EARLIER FILING DATE: 1999-08-03
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 129
;   TYPE: PRT
;   ORGANISM: Mouse
US-09-534-638-5

  Query Match             7.3%;  Score 78;  DB 4;  Length 129;
  Best Local Similarity   22.9%;  Pred. No. 0.23;
  Matches   16;  Conservative   14;  Mismatches   40;  Indels    0;  Gaps    0;

Qy      44 VALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEEALYKSKTSNKPL 103
           :|| | |   :   | ::  |:   |: |   |  | | ::|    :: |    :|  :
Db       3 IALFYILLTCSNTSPSSRNSTQSGSPQAPVVDSDRWEDTVLWAVLLQKCLLSFPSSQFSM 62

Qy     104 MIIHHLDECP 113
           :  | : |
Db      63 TPVSRLSQSP 72


RESULT 13
US-09-013-881-5
; Sequence 5, Application US/09013881
; Patent No. 6132964
;  GENERAL INFORMATION:
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Corley, Neil C.
;    APPLICANT:  Guegler, Karl J.
;    APPLICANT:  Shah, Purvi
;    TITLE OF INVENTION:  HUMAN HYDROLASE-LIKE MOLECULES
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/013,881
;      FILING DATE:  HEREWITH
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BILLINGS, LUCY J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0470 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-855-0555
;      TELEFAX:  650-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  453 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  LUNGNOT10
;      CLONE:  1376382
US-09-013-881-5

  Query Match             7.3%;  Score 78;  DB 4;  Length 453;
  Best Local Similarity   22.5%;  Pred. No. 1.5;
  Matches   45;  Conservative   34;  Mismatches   79;  Indels   42;  Gaps    9;

Qy      25 VRKSRVAMEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLI 84
           :||     :|:  :  | :::|:  |     :    | |                 :|
```

```
Db    187 LRKGGFYSQKVTTNPNLRIISLNTNLYYGPNIMTLNKTDP--------------ANQFE 231

Qy     85 WTQTYEEALYKSKTSNKPLMIIHH--LDECPHSQ---ALKKVFAE------NKEIQKLAE 133
          |    |  |  |: : : : || |  :   | ||   |::: : |       |    :|
Db    232 W---LESTLNNSQQNKEKVYIIAHVPVGYLPSSQNITAMREYYNEKLIDIFQKYSDVIAG 288

Qy    134 QF---VLLNLVYETTDKHLSPDGQYVPRIMFVDPSLTVRADITGRYSN----RLYAYEPA 186
          ||      : :   :||  ||        :|| |::|    :  : :|    ||: |:|
Db    289 QFYGHTHRDSIMVLSDKKGSPVNS-----LFVAPAVTPVKSVLEKQTNNPGIRLFQYDPR 343

Qy    187 DTALLLDNMKKALKLLKTEL 206
          |   ||| ::  | | :  |
Db    344 DYK-LLDMLQYYLNLTEANL 362


RESULT 14
PCT-US92-05401-4
; Sequence 4, Application PC/TUS9205401
;  GENERAL INFORMATION:
;    APPLICANT:  Lemischka, Ihor R.
;    TITLE OF INVENTION:  TOTIPOTENT HEMATOPOIETIC STEM CELL
;    TITLE OF INVENTION:  RECEPTORS AND THEIR LIGANDS
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  IMCLONE SYSTEMS INCORPORATED
;      STREET:  180 VARICK STREET
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  U.S.A.
;      ZIP:  10014
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/05401
;      FILING DATE:  19920626
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Feit, Irving N.
;      REGISTRATION NUMBER:  28,601
;      REFERENCE/DOCKET NUMBER:  LEM-3-PPPPPT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-645-1405
;      TELEFAX:  212-645-2054
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1160 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US92-05401-4

  Query Match             7.3%;  Score 78;  DB 5;  Length 1160;
  Best Local Similarity   28.4%;  Pred. No. 5.8;
  Matches   27;  Conservative   17;  Mismatches   35;  Indels   16;  Gaps    4;

Qy     68 RPKLPQTLSRGWGDQLIWTQTYEEALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKE 127
          ||  |  |:   | ||      |||:|::       :   : ||||:   :: |:   :
Db    952 RPSFP-NLTSFLGCQL---ADAEEAMYQN--------VDGRVSECPHTYQNRRPFSREMD 999

Qy    128 IQKLAEQFVL----LNLVYETTDKHLSPDGQYVPR 158
          :  |: | :     |||  |:  |||  |  :|:
Db   1000 LGLLSPQAQVEDSRNNLVLRTSSLHLSLTGCRLPK 1034


RESULT 15
US-08-313-185-53
; Sequence 53, Application US/08313185
; Patent No. 5851763
;  GENERAL INFORMATION:
;    APPLICANT:  Heym, Beate
;    APPLICANT:  Cole, Stewart
;    APPLICANT:  Young, Douglas
;    APPLICANT:  Zhang, Ying
;    APPLICANT:  Honore, Nadine
;    APPLICANT:  Telenti, Amalio
;    APPLICANT:  Bodmer, Thomas
;    TITLE OF INVENTION:  Rapid Detection of Antibiotic Resistance
;    TITLE OF INVENTION:  in Mycobacterium Tuberculosis
;    NUMBER OF SEQUENCES:  66
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett &
;      ADDRESSEE:  Dunner
;      STREET:  1300 I Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/313,185
;      FILING DATE:  12-OCT-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Kenneth J.
;      REGISTRATION NUMBER:  25,146
;      REFERENCE/DOCKET NUMBER:  02356.0068-00000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 408-4000
;      TELEFAX:  (202) 408-4400
;  INFORMATION FOR SEQ ID NO:  53:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  652 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-313-185-53

  Query Match             7.2%;  Score 77;  DB 2;  Length 652;
  Best Local Similarity   20.7%;  Pred. No. 3.2;
  Matches   49;  Conservative   32;  Mismatches   84;  Indels   72;  Gaps   11;

Qy     12 GPHPSRRLTQGRWVRKSRVAMEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKL 71
          ||: |  |:    :|  |: |      |   |:|  :|| :     | :     :: |
Db    211 GPNHS-PLSAAAAIRTFRMGMNDEETVA---LIAGGHTLGKTHGAGPASHVGPPEAAPIE 266

Qy     72 PQTLSRGWGD-------------QLIWTQT----------YEEALYKS---------- 96
          | |  ||                :::|||           ||  | ||
Db    267 AQGL--GWASSYGSGVGADAITSGEVVWTQTPTQWNFFENLFYEWVLTKSPAGAQEAVDG 324

Qy     97 ----------KTSNKPLMIIHHLDEC-PHSQALKKVFAENKEIQKLAEQFVLLNLVYET 144
                    || |::  |    |  : : : |  : | :  |  :      ::
Db    325 APDIIPDPFDPSKKRKPTMLVTDLLRFDPEYEKISRRFLNDPEFEAFARAW------FKL 378

Qy    145 TDKHLSPD---GQYVPR--IMFVDPSLTVRADI--------TGRYSNRLYAYEPADT 188
          | : : |    |  ||:  ::: ||      ||        :|  |  : |:  | |
Db    379 THRDMGPKRYIGPEVPKEDLIWQDPPQYPTEDIILKAAIAASGLVSELVSAWASAST 435


Search completed: October  9, 2002, 16:41:41
Job time : 19 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 16:40:27 ; Search time 32 Seconds
(without alignments)
715.038 Million cell updates/sec

Title: US-09-674-266A-181
Perfect score: 206
Sequence: 1 RLSCAGTLSGSGPHPSRRLT.....DTALLLDNMKKALKLLKTEL 206

Scoring table: [redacted]OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 111073796 residues

Word size : 0

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	206	100.0	206	20	AAY73837	Human prostate tum
2	194	94.2	194	20	AAY76590	Human ovarian tumo
3	194	94.2	202	22	AAG75621	Human colon cancer
4	175	85.0	175	19	AAW77365	zsig10 polypeptide
5	175	85.0	175	19	AAW37844	Human XAG growth f
6	175	85.0	175	19	AAW37872	Human protein comp
7	175	85.0	175	20	AAY59675	Secreted protein 1
8	175	85.0	175	21	AAB24070	Human PRO1030 prot
9	175	85.0	175	22	AAU08804	Breast cancer cell
10	175	85.0	175	22	AAB72203	Human huXAG-1/CCSG
11	126	61.2	132	21	AAG00109	Human secreted pro
12	116	56.3	116	20	AAY12312	Human 5' EST secre
13	115	55.8	115	21	AAY64672	Human 5' EST relat
14	59	28.6	131	20	AAY59718	Secreted protein 7
15	55	26.7	56	20	AAY11881	Human 5' EST secre
16	42	20.4	62	20	AAY11882	Human 5' EST secre
17	29	14.1	89	20	AAY76568	Human ovarian tumo
18	24	11.7	166	19	AAW37846	Human XAG growth f
19	24	11.7	166	21	AAB00194	Breast cancer prot
20	24	11.7	166	22	AAU25728	Breast cancer-asso
21	24	11.7	166	22	AAB72205	Human huXAG-3/CCSG
22	24	11.7	166	22	AAB31192	Amino acid sequenc
23	24	11.7	166	22	AAU07647	Human BCMP 11 poly
24	24	11.7	168	22	AAM24502	Colon tumour relat
25	17	8.3	56	20	AAY11939	Human 5' EST secre
26	14	6.8	39	21	AAB58921	Breast and ovarian
27	11	5.3	11	22	AAU08805	Breast cancer cell
28	11	5.3	11	22	AAU08806	Breast cancer cell
29	11	5.3	11	22	AAU25804	Breast cancer-asso
30	11	5.3	11	22	AAU25805	Breast cancer-asso
31	10	4.9	11	22	AAU25806	Breast cancer-asso
32	10	4.9	70	20	AAY11654	Human 5' EST secre
33	10	4.9	180	22	AAG73737	Human colon cancer
34	9	4.4	9	22	AAU25812	Breast cancer-asso
35	9	4.4	9	22	AAU07648	Human BCMP 11 pept
36	8	3.9	315	22	AAG73015	Olfactory receptor
37	8	3.9	1284	22	ABB61690	Drosophila melanog
38	7	3.4	10	22	AAM43189	Mycoplasma genital
39	7	3.4	73	22	AAM86554	Human immune/haema
40	7	3.4	105	22	AAM91788	Human immune/haema
41	7	3.4	105	22	AAO04178	Human polypeptide
42	7	3.4	109	21	AAG03933	Human secreted pro
43	7	3.4	109	22	AAM40426	Human polypeptide
44	7	3.4	109	22	AAB80983	Human hARE-RBP1.
45	7	3.4	119	22	AAM25807	Human protein sequ

ALIGNMENTS

RESULT 1
AAY73837
ID AAY73837 standard; Protein; 206 AA.
XX
AC AAY73837;
XX
DT 14-MAR-2000 (first entry)
XX
DE Human prostate tumor EST fragment derived protein #24.
XX
KW Pancreas; tumor; EST; expressed sequence tag; human; cytostatic;
KW treatment.
XX
OS Homo sapiens.
XX
PN DE19820190-A1.
XX
PD 04-NOV-1999.
XX
PF 28-APR-1998; 98DE-1020190.
XX
PR 28-APR-1998; 98DE-1020190.
XX
PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
XX
DR WPI; 1999-621386/54.
DR N-PSDB; AAZ52865.
XX
PT New human nucleic acid sequences from pancreatic tumors, and related
PT proteins -
XX
PS Claim 23; Page 318; 502pp; German.

```
XX
CC   This invention describes novel polypeptides and their encoding nucleic
CC   acids derived from human pancreatic tumor tissue which have cytostatic
CC   activity. The sequences are also useful in producing pharmaceutical
CC   compositions for treatment of pancreatic tumors. AAY73814-Y74252
CC   represent protein fragments encoded by the human pancreatic tumor cDNA
CC   library derived expressed sequence tag (EST) sequences represented in
CC   AAZ52858-Z53014.
XX
SQ   Sequence   206 AA;

  Query Match            100.0%;  Score 206;  DB 20;  Length 206;
  Best Local Similarity  100.0%;  Pred. No. 5e-194;
  Matches  206;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 RLSCAGTLSGSGPHPSRRLTQGRWVRKSRVAMEKIPVSAFLLLVALSYTLARDTTVKPGA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 RLSCAGTLSGSGPHPSRRLTQGRWVRKSRVAMEKIPVSAFLLLVALSYTLARDTTVKPGA 60

Qy     61 KKDTKDSRPKLPQTLSRGWGDQLIWTQTYEEALYKSKTSNKPLMIIHHLDECPHSQALKK 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 KKDTKDSRPKLPQTLSRGWGDQLIWTQTYEEALYKSKTSNKPLMIIHHLDECPHSQALKK 120

Qy    121 VFAENKEIQKLAEQFVLLNLVYETTDKHLSPDGQYVPRIMFVDPSLTVRADITGRYSNRL 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 VFAENKEIQKLAEQFVLLNLVYETTDKHLSPDGQYVPRIMFVDPSLTVRADITGRYSNRL 180

Qy    181 YAYEPADTALLLDNMKKALKLLKTEL 206
          |||||||||||||||||||||||||||
Db    181 YAYEPADTALLLDNMKKALKLLKTEL 206


RESULT 2
AAY76590
ID   AAY76590 standard; Protein; 194 AA.
XX
AC   AAY76590;
XX
DT   10-APR-2000  (first entry)
XX
DE   Human ovarian tumor EST fragment encoded protein 86.
XX
KW   Expressed sequence tag; EST; human; ovarian tumor; anticancer;
KW   gene therapy; treatment.
XX
OS   Homo sapiens.
XX
PN   DE19817557-A1.
XX
PD   21-OCT-1999.
XX
PF   09-APR-1998;   98DE-1017557.
XX
PR   09-APR-1998;   98DE-1017557.
XX
PA   (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI   Rosenthal A,  Specht T,  Hinzmann B,  Schmitt A,  Pilarsky C,  Dahl E;
XX
DR   WPI; 1999-591920/51.
DR   N-PSDB; AAZ77486.
XX
PT   New nucleic acid sequences expressed in ovarian, and some other, cancer
PT   tissues, and derived polypeptides, for treatment of ovarian cancer and
PT   identification of therapeutic agents  -
XX
PS   Claim 25; Page 279; 310pp; German.
XX
CC   This invention describes novel nucleic acid (cDNA) sequences (A) which
CC   have anticancer activity and are highly expressed in ovarian tumor
CC   tissue (and some also in testis and breast cancer tissue). The products
CC   of the invention can be used for gene therapy. (A) are used (i) for
CC   recombinant expression of polypeptides (B) and (ii) to isolate complete
CC   genes. (B) are used (i) to identify agents suitable for treatment of
CC   ovarian cancer; (ii) directly for treating this form of cancer
CC   (including expression from gene therapy vectors) and (iii) for generation
CC   of specific antibodies. (A) are identified by assembling ESTs (expressed
CC   sequence tags) from a particular tissue type before comparison of
CC   expression patterns. This allows a significantly longer fragment of the
CC   gene to be revealed, so should reduce the number of failures associated
CC   with the fact that ESTs from different libraries may represent different
CC   parts of the same unknown gene, distorting the estimated frequency of
CC   occurrence in a particular tissue. AAY76505-Y76638 represent protein
CC   fragments encoded by the human ovarian tumor cDNA library derived EST
CC   fragments represented in AAZ77450-Z77572.
XX
SQ   Sequence   194 AA;

  Query Match            94.2%;  Score 194;  DB 20;  Length 194;
  Best Local Similarity  100.0%;  Pred. No. 2.9e-182;
  Matches  194;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     13 PHPSRRLTQGRWVRKSRVAMEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLP 72
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 PHPSRRLTQGRWVRKSRVAMEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLP 60

Qy     73 QTLSRGWGDQLIWTQTYEEALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLA 132
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 QTLSRGWGDQLIWTQTYEEALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLA 120

Qy    133 EQFVLLNLVYETTDKHLSPDGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLL 192
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 EQFVLLNLVYETTDKHLSPDGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLL 180

Qy    193 DNMKKALKLLKTEL 206
          ||||||||||||||
Db    181 DNMKKALKLLKTEL 194


RESULT 3
AAG75621
ID   AAG75621 standard; Protein; 202 AA.
XX
AC   AAG75621;
XX
DT   03-SEP-2001  (first entry)
XX
DE   Human colon cancer antigen protein SEQ ID NO:6385.
XX
KW   Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW   colorectal carcinoma.
XX
OS   Homo sapiens.
XX
PN   WO200122920-A2.
XX
PD   05-APR-2001.
XX
PF   28-SEP-2000; 2000WO-US26524.
XX
PR   29-SEP-1999;   99US-0157137.
PR   03-NOV-1999;   99US-0163280.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Barash SC,  Birse CE,  Rosen CA;
XX
DR   WPI; 2001-235357/24.
DR   N-PSDB; AAH35026.
XX
PT   Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT   useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS   Claim 11; Page 7865-7866; 9803pp; English.
XX
CC   AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
```

```
CC   cancer-associated nucleic acid molecules (N) and proteins (P), where
CC   the proteins are collectively known as colon cancer antigens. The colon
CC   cancer antigens have cytostatic activity and can be used in gene
CC   therapy and vaccine production. N and P may be used in the prevention,
CC   diagnosis and treatment of diseases associated with inappropriate P
CC   expression. For example, N and P may be used to treat disorders
CC   associated with decreased expression by rectifying mutations or deletions
CC   in a patient's genome that affect the activity of P by expressing
CC   inactive proteins or to supplement the patients own production of P.
CC   Additionally, N may be used to produce the colon cancer-associated Ps,
CC   by inserting the nucleic acids into a host cell and culturing the cell
CC   to express the proteins. N and P can be used in the prevention, diagnosis
CC   and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC   and AAB77789 represent sequences used in the exemplification of the
CC   present invention.
CC   N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC   missing at time of publication, meaning no sequences are present for
CC   SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ   Sequence   202 AA;

  Query Match             94.2%;  Score 194;  DB 22;  Length 202;
  Best Local Similarity   100.0%;  Pred. No. 3e-182;
  Matches  194;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      13 PHPSRRLTQGRWVRKSRVAMEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLP 72
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9 PHPSRRLTQGRWVRKSRVAMEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLP 68

Qy      73 QTLSRGWGDQLIWTQTYEEALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLA 132
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69 QTLSRGWGDQLIWTQTYEEALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLA 128

Qy     133 EQFVLLNLVYETTDKHLSPDGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLL 192
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     129 EQFVLLNLVYETTDKHLSPDGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLL 188

Qy     193 DNMKKALKLLKTEL 206
           ||||||||||||||
Db     189 DNMKKALKLLKTEL 202


RESULT 4
AAW77365
ID   AAW77365 standard; Protein; 175 AA.
XX
AC   AAW77365;
XX
DT   21-DEC-1998  (first entry)
XX
DE   zsig10 polypeptide.
XX
KW   Human; mucous-mediated function; adhesion; tumour metastasis;
KW   bacterial colonisation; microbial infection; AIDS; cystic fibrosis;
KW   chronic obstructive pulmonary disease; asthma; Crohn's disease;
KW   sinonasal inflammatory disease; inflammatory bowel disease; bronchitis.
XX
OS   Homo sapiens.
XX
PN   WO9841627-A1.
XX
PD   24-SEP-1998.
XX
PF   18-MAR-1998;   98WO-US05251.
XX
PR   19-MAR-1997;   97US-0039631.
XX
PA   (ZYMO ) ZYMOGENETICS INC.
XX
PI   Sheppard PO;
XX
DR   WPI; 1998-531566/45.
DR   N-PSDB; AAV59320.
XX
PT   New isolated mucous-associated polypeptide, zsig10 - used to develop
PT   products for treating e.g. tumour metastasis, microbial infections,
PT   cystic fibrosis, asthma, bronchitis or inflammatory bowel disease
XX
PS   Claim 1; Page 82; 109pp; English.
XX
CC   The human polypeptide zsig10 is involved in mucous-mediated functions
CC   such as adhesion.  The products of the invention can be used in the study
CC   and treatment of e.g. tumour metastasis, bacterial colonisation,
CC   susceptibility to and persistence of infection, microbial infections,
CC   AIDS, cystic fibrosis, chronic obstructive pulmonary disease, asthma,
CC   sinonasal inflammatory disease, inflammatory bowel disease, bronchitis,
CC   or Crohn's disease.  The products can also be used for detection,
CC   diagnosis and drug screening.
XX
SQ   Sequence   175 AA;

  Query Match             85.0%;  Score 175;  DB 19;  Length 175;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-163;
  Matches  175;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      32 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 91
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 60

Qy      92 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 151
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 120

Qy     152 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 206
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 175


RESULT 5
AAW37844
ID   AAW37844 standard; Protein; 175 AA.
XX
AC   AAW37844;
XX
DT   28-AUG-1998  (first entry)
XX
DE   Human XAG growth factor huXAG-1.
XX
KW   HuXAG-1; XAG; growth factor; colon cancer; tumour marker;
KW   breast disease; liver disease; lung disease; emphysema;
KW   wound healing; diagnosis; therapy; human.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..20
FT                   /label= Sig_peptide
FT   Protein         21..175
FT                   /label= Mat_protein
FT                   /note= "Claim 11"
FT   Peptide         20..43
FT                   /label= Epitope
FT                   /note= "Claim 12"
FT   Peptide         44..52
FT                   /label= Epitope
FT                   /note= "Claim 12"
FT   Peptide         61..72
FT                   /label= Epitope
FT                   /note= "Claim 12"
FT   Peptide         90..103
FT                   /label= Epitope
FT                   /note= "Claim 12"
FT   Peptide         113..125
FT                   /label= Epitope
FT                   /note= "Claim 12"
FT   Peptide         138..150
```

```
FT                   /label= Epitope
FT                   /note= "Claim 12"
XX
PN   WO9807749-A1.
XX
PD   26-FEB-1998.
XX
PF   22-AUG-1997;   97WO-US14139.
XX
PR   23-AUG-1996;   96WO-US13766.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Dillon PJ,  Ebner R,  Endress GA,  Yu G;
XX
DR   WPI; 1998-169093/15.
DR   N-PSDB; AAV19155.
XX
PT   New isolated human XAG growth factor(s) - used to develop products
PT   for treating e.g. liver, lung or breast diseases or
PT   hyperproliferative disorders, e.g. cancer.
XX
PS   Claim 1; Fig 1; 141pp; English.
XX
CC   This polypeptide comprises huXAG-1, a member of a novel family of
CC   human growth factors also including huXAG-2 (see AAW37845) and
CC   huXAG-3 (see AAW37846).  These proteins share homology with the
CC   XAG protein of Xenopus laevis, which is involved in embryogenesis
CC   and is expressed in adult tissue.  huXAG-1 is specifically found
CC   in cancerous colon cells and may therefore be a growth factor for
CC   colon cancer.  huXAG-1 cDNA (see AAV19155) was isolated from a cDNA
CC   library derived from human colon cancer tissue.  Vectors, host
CC   cells, antibodies, and screening methods for identifying agonists
CC   and antagonists of huXAG-1 are provided.  HuXAG polypeptides are
CC   growth factors and can be used to stimulate proliferation of cells.
CC   They can be used to stimulate the proliferation and differentiation
CC   of hepatocytes to alleviate or treat liver diseases and pathologies
CC   such as fulminant liver failure caused by cirrhosis, liver damage
CC   caused by viral hepatitis and toxic substances.  They can also be
CC   used to stimulate or promote liver regeneration, e.g. after
CC   surgery.  They can also be used to prevent and heal damage to the
CC   lungs caused by various pathological states.  They can be used to
CC   stimulate proliferation and differentiation and promote the repair
CC   of alveoli and bronchiolar epithelium to prevent, attenuate, or
CC   treat acute or chronic lung damage, e.g. emphysema, which results
CC   in the progressive loss of alveoli, and inhalation injuries, e.g.
CC   resulting from smoke inhalation and burns, that cause necrosis of
CC   the bronchiolar epithelium and alveoli.  They can also be used to
CC   stimulate the proliferation and differentiation of breast tissue
CC   and could therefore be used to promote healing of breast tissue
CC   injury due to surgery, trauma or cancer.  Antagonists can be used
CC   to treat hyperproliferative disorders, including cancer, in
CC   particular hepatocellular carcinoma, osteoclastoma, breast cancer,
CC   or colon cancer.  The products can also be used for detection and
CC   diagnosis.
XX
SQ   Sequence   175 AA;

  Query Match             85.0%;  Score 175;  DB 19;  Length 175;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-163;
  Matches  175;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      32 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 91
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 60

Qy      92 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 151
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 120

Qy     152 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 206
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 175


RESULT 6
AAW37872
ID   AAW37872 standard; Protein; 175 AA.
XX
AC   AAW37872;
XX
DT   10-AUG-1998  (first entry)
XX
DE   Human protein comprising secretory signal amino acid sequence 9.
XX
KW   Human protein; secretory signal; nutritional source; cytokine;
KW   immunity;haematopoiesis; activin; inhibin; tumour; chemotactic;
KW   chemokinetic; thrombolytic; anti-inflammatory; inhibition;
KW   stomach cancer cell.
XX
OS   Homo sapiens.
XX
XX
PN   WO9811217-A2.
XX
PD   19-MAR-1998.
XX
PF   12-SEP-1997;   97WO-JP03239.
XX
PR   13-SEP-1996;   96JP-0243060.
XX
PA   (PROT-) PROTEGENE INC.
PA   (SAGA ) SAGAMI CHEM RES CENTRE.
XX
PI   Kato S,  Kobayashi M,  Sekine S,  Yamaguchi T;
XX
DR   WPI; 1998-207380/18.
DR   N-PSDB; AAV29047, AAV29048.
XX
PT   Human proteins with secretory signal sequences - used to treat
PT   immune deficiencies, infections, tumours, and haematopoietic
PT   disorders, etc.
XX
PS   Claim 1; Pages 79; 131pp; English.
XX
CC   This is the amino acid sequence of a novel human protein comprising
CC   a secretory signal isolated from stomach cancer cells.  Its proteins
CC   can be used as nutritional sources or supplements.  The proteins may
CC   also have cytokine functions, immune modulating functions,
CC   haematopoiesis regulating activity, activin/inhibin regulating
CC   activity, chemotactic/chemokinetic activity, haemostatic and
CC   thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC   activity, tumour inhibition activity.
XX
SQ   Sequence   175 AA;

  Query Match             85.0%;  Score 175;  DB 19;  Length 175;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-163;
  Matches  175;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      32 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 91
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 60

Qy      92 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 151
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 120

Qy     152 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 206
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 175


RESULT 7
AAY59675
ID   AAY59675 standard; Protein; 175 AA.
```

```
XX
AC   AAY59675;
XX
DT   18-JAN-2000  (first entry)
XX
DE   Secreted protein 108-008-5-0-A6-FL.
XX
KW   Secreted protein; fingerprint identification technique;
KW   chromosome mapping; human; hereditary disease; diagnosis; cancer;
KW   hyperlipidaemia; cardiovascular; neurodegenerative disorder; therapy;
KW   autoimmune disease; rheumatic disease; embryogenic disorder; myopathy;
KW   renal injury; amino aciduria; hypoglycaemia; male rat infertility;
KW   hypertension.
XX
OS   Homo sapiens.
XX
PN   WO9940189-A2.
XX
PD   12-AUG-1999.
XX
PF   09-FEB-1999;   99WO-IB00282.
XX
PR   09-FEB-1998;   98US-0074121.
PR   13-APR-1998;   98US-0081563.
PR   10-AUG-1998;   98US-0096116.
PR   04-SEP-1998;   98US-0099273.
XX
PA   (GEST ) GENSET.
XX
PI   Bougueleret L,  Duclert A,  Dumas Milne Edwards J;
XX
DR   WPI; 1999-600966/51.
DR   N-PSDB; AAZ40803.
XX
PT   Extended cDNAs useful for expressing secreted proteins and to obtain
PT   specific antibodies -
XX
PS   Claim 10; Page 199; 244pp; English.
XX
CC   This sequence represents a human secreted protein of the invention.
CC   The extended cDNAs (or genomic DNAs obtainable from them) may be used to
CC   prepare PCR primers and probes. These are useful for forensic matching or
CC   positive identification by DNA sequencing. They may also be used in
CC   alternative fingerprint identification techniques. Antibodies against the
CC   proteins encoded by the extended cDNAs are useful in identification of
CC   tissue types or cell species, as well as identifying tissue specific
CC   soluble proteins. The sequences can be used for chromosome mapping and
CC   identification of genes associated with hereditary diseases or drug
CC   response. signal sequences from the cDNAs can be used in construction of
CC   secretion vectors. Other sequences derived from the extended cDNAs can be
CC   used to clone upstream genomic DNA sequences including promoters. This is
CC   in turn useful for identifying proteins that interact with promoter
CC   sequences. Some of the proteins may be useful in diagnosing and treating
CC   several disorders including, but not limited to: cancer, hyperlipidaemia,
CC   cardiovascular and neurodegenerative disorders, autoimmune diseases, and
CC   rheumatic diseases, embryogenic disorders, hypertension, renal injury,
CC   amino acidurias, hypoglycaemia, male rat infertility and myopathies.
XX
SQ   Sequence   175 AA;

  Query Match             85.0%;  Score 175;  DB 20;  Length 175;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-163;
  Matches  175;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      32 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 91
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 60

Qy      92 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 151
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 120

Qy     152 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 206
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 175


RESULT 8
AAB24070
ID   AAB24070 standard; Protein; 175 AA.
XX
AC   AAB24070;
XX
DT   29-JAN-2001  (first entry)
XX
DE   Human PRO1030 protein sequence SEQ ID NO:40.
XX
KW   Human; tumour; diagnosis; neoplastic disease; neoplastic cell growth;
KW   proliferation; tumourigenesis; identification; cancer; cytostatic;
KW   nootropic; neuroprotective; antiinflammatory; immunosuppressive;
KW   immunostimulant; antiangiogenic; leukaemia; lymphoid malignancy;
KW   neuronal disorder; glial disorder; astrocytal disorder; angiogenic;
KW   hypothalamic disorder; glandular disorder; macrophagal disorder;
KW   epithelial disorder; stromal disorder; blastocoelic disorder;
KW   inflammatory disorder; immunologic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200053755-A2.
XX
PD   14-SEP-2000.
XX
PF   06-JAN-2000; 2000WO-US00376.
XX
PR   08-MAR-1999;   99WO-US05028.
PR   02-JUN-1999;   99WO-US12252.
PR   23-JUN-1999;   99US-0141037.
PR   07-JUL-1999;   99US-0143048.
PR   26-JUL-1999;   99US-0145698.
PR   30-NOV-1999;   99WO-US28313.
PR   20-DEC-1999;   99WO-US30911.
PR   05-JAN-2000; 2000WO-US00219.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Ashkenazi AJ,  Baker KP,  Goddard A,  Gurney AL,  Hillan KJ,  Roy MA;
PI   Watanabe CK,  Wood WI;
XX
DR   WPI; 2000-572270/53.
DR   N-PSDB; AAC58380.
XX
PT   Thirty PRO polynucleotides encoding PRO polypeptides, useful in the
PT   treatment, diagnosis and prevention of cancer -
XX
PS   Claim 61; Fig 28; 286pp; English.
XX
CC   The present invention describes an isolated antibody  that binds to
CC   one of the human PRO proteins designated PRO212, PRO290, PRO341, PRO535,
CC   PRO619, PRO717, PRO809, PRO830, PRO848, PRO943, PRO1005, PRO1009,
CC   PRO1025, PRO1030, PRO1097, PRO1107, PRO1111, PRO1153, PRO1182, PRO1184,
CC   PRO1187, PRO1281, PRO23, PRO39, PRO834, PRO1317, PRO1710, PRO2094,
CC   PRO2145 OR PRO2198. PRO antagonists can be used to inhibit tumour cell
CC   growth. The PRO polypeptides and nucleotides are useful in the
CC   treatment, diagnosis and prevention of cancer. The antibodies and other
CC   anti-tumour compounds maybe used to treat various conditions, including
CC   those characterised by overexpression and/or activation of the amplified
CC   PRO genes. Exemplary conditions or disorders to be treated with such
CC   antibodies and other compounds include benign or malignant tumours
CC   (e.g., renal, liver, kidney, bladder, breast, gastric, ovarian,
CC   colorectal, prostate, pancreatic, lung, vulva, thyroid, hepatic
CC   carcinomas, sarcomas, glioblastomas, and various head and neck tumours),
CC   leukaemias and lymphoid malignancies, other disorders such as neuronal,
CC   glial, astrocytal, hypothalamic and other glandular, macrophagal,
CC   epithelial, stromal and blastocoelic disorders, and inflammatory,
CC   angiogenic and immunologic disorders. AAC58242 to AAC58366 represent PCR
CC   primers and hybridisation probes used in the isolation of the human PRO
```

```
CC   sequences. AAC58367 to AAC58396 and AAB24057 to AAB24089 represent human
CC   PRO polynucleotide and protein sequences given in the exemplification of
CC   the present invention.
XX
SQ   Sequence   175 AA;

  Query Match             85.0%;  Score 175;  DB 21;  Length 175;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-163;
  Matches  175;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      32 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 91
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 60

Qy      92 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 151
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 120

Qy     152 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 206
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 175


RESULT 9
AAU08804
ID   AAU08804 standard; Protein; 175 AA.
XX
AC   AAU08804;
XX
DT   19-DEC-2001  (first entry)
XX
DE   Breast cancer cell membrane associated protein 7 (BCMP 7).
XX
KW   Breast cancer cell membrane protein 7; BCMP 7; breast cancer;
KW   metastasis; cytostatic; gene therapy; antibody; antisense;
KW   chromosome 7p21.3; vaccine.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..20
FT                   /note= "Signal peptide"
FT   Protein         21..175
FT                   /note= "Mature BCMP 7"
XX
PN   WO200163290-A1.
XX
PD   30-AUG-2001.
XX
PF   21-FEB-2001; 2001WO-GB00734.
XX
PR   25-FEB-2000; 2000GB-0004576.
XX
PA   (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI   Boyd RS,  Stamps AC,  Terrett JA,  Tyson KL;
XX
DR   WPI; 2001-570651/64.
DR   N-PSDB; AAS13480.
XX
PT   Diagnosing, preventing and treating breast cancer using a breast cancer
PT   cell membrane protein BCMP 7 -
XX
PS   Claim 1; Fig 1; 62pp; English.
XX
CC   The invention describes the novel use of a protein found in breast
CC   cancer cell membranes (BCMP 7) for diagnosing, preventing and treating
CC   breast cancers. The peptide has cytostatic action and potential uses in
CC   gene therapy and in vaccines. The polypeptide, antisense nucleic acids,
CC   or fusion proteins comprising and Green Fluorescent protein or the DsRed
CC   Fluorescent protein, antibodies specific for and/or nucleic acid are
CC   used for the prevention and/or treatment of breast cancer. Antibodies
CC   against may also be used for screening for and/or diagnosis of breast
CC   cancer in a patient. The method for monitoring/assessing breast cancer
CC   treatment in a patient and for the identification of metastatic breast
CC   cancer cells in samples from a patient. This sequence is breast
CC   cancer cell associated protein 7 (BCMP 7), encoded by a gene located on
CC   chromosome 7p21.3, described in the method of the invention.
XX
SQ   Sequence   175 AA;

  Query Match             85.0%;  Score 175;  DB 22;  Length 175;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-163;
  Matches  175;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      32 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 91
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 60

Qy      92 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 151
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 120

Qy     152 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 206
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 175


RESULT 10
AAB72203
ID   AAB72203 standard; Protein; 175 AA.
XX
AC   AAB72203;
XX
DT   04-MAY-2001  (first entry)
XX
DE   Human huXAG-1/CCSG colon cancer specific gene amino acid sequence.
XX
KW   Human; growth factor; huXAG-1; colon cancer specific gene; CCSG;
KW   cell proliferation; liver disease; fulminant liver failure; cirrhosis;
KW   hepatitis; cancer; colon cancer; colorectal carcinoma.
XX
OS   Homo sapiens.
XX
PN   US6171816-B1.
XX
PD   09-JAN-2001.
XX
PF   22-AUG-1997;   97US-0916576.
XX
PR   23-AUG-1996;   96US-0024347.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Yu G,  Dillon PJ,  Ebner R,  Endress GA;
XX
DR   WPI; 2001-136872/14.
DR   N-PSDB; AAF63314.
XX
PT   Novel human growth factor polypeptide useful for diagnosing and
PT   treating colon cancer and liver diseases, to prevent and heal damage to
PT   the lungs and for identifying modulators of therapeutic use -
XX
PS   Claim 165; Fig 1; 59pp; English.
XX
CC   This invention relates to a human growth factor polypeptide huXAG-1 also
CC   known as a colon cancer specific gene (CCSG). HuXAG-1 stimulates cell
CC   proliferation as a growth factor. The HuXAG-1 protein is useful for
CC   identifying compounds capable of enhancing or inhibiting cellular
CC   response induced by huXAG-1. The protein is also useful for stimulating
CC   proliferation of cells e.g. colon, breast, liver and lung cells, and
CC   hepatocytes. It is useful for alleviating or treating liver diseases and
CC   pathologies such as fulminant liver failure caused by cirrhosis, liver
CC   damage caused by viral hepatitis and toxic substances, for preventing and
CC   treating damage to the lungs caused by various pathological states and
CC   for promoting healing of breast tissue injury due to surgery, trauma or
```

```
CC   cancer. HuXAG-1 and the identified antagonist are useful for treating
CC   cancer, in particular colon cancer. Detecting altered levels of huXAG-1
CC   and its polynucleotides are useful for diagnosing or detecting cancer in
CC   mammals. The gene encoding huXAG-1 is useful for monitoring human
CC   colorectal carcinoma. huXAG-1 nucleic acid molecules are also useful for
CC   chromosome identification. The present sequence represents the huXAG-1
CC   protein.
XX
SQ   Sequence   175 AA;

  Query Match             85.0%;  Score 175;  DB 22;  Length 175;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-163;
  Matches  175;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      32 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 91
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 60

Qy      92 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 151
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 120

Qy     152 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 206
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 175


RESULT 11
AAG00109
ID   AAG00109 standard; Protein; 132 AA.
XX
AC   AAG00109;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein, SEQ ID NO: 4190.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   N-PSDB; AAC00115.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 13; SEQ ID 4190; 71pp + CD-ROM; English.
XX
CC   The present sequence is a polypeptide encoded by one of a large number
CC   of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC   were prepared from total human RNAs or polyA+ RNAs derived from 30
CC   different tissues. EST sequences usually correspond mainly to the 3'
CC   untranslated region (UTR) of the mRNA because they are often obtained
CC   from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC   isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC   those cases where longer cDNA sequences have been obtained, the full 5'
CC   UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC   ends and can therefore be used to obtain full length cDNAs and genomic
CC   DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC   chromosome mapping procedures. They are used to obtain upstream
CC   regulatory sequences and to design expression and secretion vectors.
XX
SQ   Sequence   132 AA;

  Query Match             61.2%;  Score 126;  DB 21;  Length 132;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-115;
  Matches  126;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      32 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 91
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 60

Qy      92 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 151
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 120

Qy     152 DGQYVP 157
           ||||||
Db     121 DGQYVP 126


RESULT 12
AAY12312
ID   AAY12312 standard; Protein; 116 AA.
XX
AC   AAY12312;
XX
DT   17-JUN-1999  (first entry)
XX
DE   Human 5' EST secreted protein SEQ ID NO:343.
XX
KW   Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW   forensic; gene therapy; chromosome mapping; signal peptide;
KW   upstream regulatory sequence; cytokine activity; cell proliferation;
KW   differentiation; haematopoiesis regulation; tissue growth regulation;
KW   reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; anti-inflammatory; tumour inhibition.
XX
OS   Homo sapiens.
XX
PN   WO9906548-A2.
XX
PD   11-FEB-1999.
XX
PF   31-JUL-1998;   98WO-IB01222.
XX
PR   01-AUG-1997;   97US-0905135.
XX
PA   (GEST ) GENSET.
XX
PI   Duclert A,  Dumas Milne Edwards J,  Lacroix B;
XX
DR   WPI; 1999-153778/13.
DR   N-PSDB; AAX41145.
XX
PT   New nucleic acids encoding human secreted proteins - obtained from
PT   cDNA libraries prepared from e.g. liver, ovary, brain, prostate,
PT   kidney, lung, umbilical cord, placenta and colon tissue
XX
PS   Claim 27; Page 682; 824pp; English.
XX
CC   AAX41094 to AAX41347 represent 5' expressed sequence tags (ESTs) for
CC   human secreted proteins, and encode the proteins given in AAY12261 to
CC   AAY12514, respectively. The proteins given represent the signal peptide
CC   and an N-terminal fragment of a secreted protein. The nucleic acid
CC   sequences can be used for producing secreted human gene products. They
CC   can also be used to develop products for diagnosis and therapy. The
CC   proteins obtained may have cytokine activity, cell
CC   proliferation/differentiation activity, haematopoiesis regulating
CC   activity, tissue growth regulating activity, reproductive hormone
CC   regulating activity, chemotactic/ chemokinetic activity, haemostatic and
```

```
CC   thrombolytic activity, receptor/ ligand activity, anti-inflammatory
CC   activity, tumour inhibition activity or other activities. The products
CC   can be used in forensic, gene therapy and chromosome mapping procedures.
CC   The sequences can also be used for obtaining corresponding promoter
CC   sequences. The nucleic acids encoding the signal peptide can be used for
CC   directing extracellular secretion of a polypeptide or the insertion of a
CC   polypeptide into a membrane, or importing a polypeptide into a cell.
XX
SQ   Sequence   116 AA;

  Query Match             56.3%;  Score 116;  DB 20;  Length 116;
  Best Local Similarity   100.0%;  Pred. No. 6.7e-106;
  Matches  116;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      32 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 91
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 60

Qy      92 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDK 147
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDK 116


RESULT 13
AAY64672
ID   AAY64672 standard; Protein; 115 AA.
XX
AC   AAY64672;
XX
DT   01-FEB-2000  (first entry)
XX
DE   Human 5' EST related polypeptide SEQ ID NO:833.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
KW   gene therapy; chromosome mapping; upstream regulatory sequence;
KW   forensic; location; development; protein synthesis; stability;
KW   regulation; identification.
XX
OS   Homo sapiens.
XX
PN   WO9953051-A2.
XX
PD   21-OCT-1999.
XX
PF   09-APR-1999;   99WO-IB00712.
XX
PR   09-APR-1998;   98US-0057719.
PR   28-APR-1998;   98US-0069047.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-038446/03.
DR   N-PSDB; AAZ42286.
XX
PT   Novel secreted protein 5' expressed sequence tag sequences used in
PT   diagnostic, forensic, gene therapy, and chromosome mapping procedures
XX
PS   Claim 3; Page 604; 837pp; English.
XX
CC   AAZ42265 to AAZ43075 represent novel 5' expressed sequence tag (EST)
CC   sequences, corresponding to human secreted proteins. AAY64651 to
CC   AAY65438 represent the EST-related proteins corresponding to AAZ42265 to
CC   AAZ43052. The 5' ESTs can be used for producing secreted human gene
CC   products. They can be used to identify and isolate 5' untranslated
CC   regions (UTRs) and upstream regulatory regions which control the
CC   location, development stage, rate, and quantity of protein synthesis, as
CC   well as stability of mRNA. The ESTs are also useful as probes for
CC   chromosome mapping, and to obtain full length cDNA clones. The ESTs can
CC   also be used in forensic procedures to identify individuals, or in
CC   diagnostic procedures to identify individuals having genetic diseases
CC   resulting from abnormal gene expression. The products may also be used in
CC   gene therapy protocols. The nucleic acids encoding signal peptides can be
CC   used for directing extracellular secretion of a polypeptide or the
CC   insertion of a polypeptide into a membrane, or importing a polypeptide
CC   into a cell. The proteins encoded by the EST sequences may be useful in
CC   treating a variety of human conditions. Secreted proteins have
CC   therapeutic value, and the identification of new secreted proteins is
CC   valuable. AAZ42249 to AAZ42264 and AAY64644 to AAY64650 represent
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence   115 AA;

  Query Match             55.8%;  Score 115;  DB 21;  Length 115;
  Best Local Similarity   100.0%;  Pred. No. 6.4e-105;
  Matches  115;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      32 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 91
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 60

Qy      92 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTD 146
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTD 115


RESULT 14
AAY59718
ID   AAY59718 standard; Protein; 131 AA.
XX
AC   AAY59718;
XX
DT   18-JAN-2000  (first entry)
XX
DE   Secreted protein 78-21-1-B7-FL1.
XX
KW   Secreted protein; fingerprint identification technique;
KW   chromosome mapping; human; hereditary disease; diagnosis; cancer;
KW   hyperlipidaemia; cardiovascular; neurodegenerative disorder; therapy;
KW   autoimmune disease; rheumatic disease; embryogenic disorder; myopathy;
KW   renal injury; amino aciduria; hypoglycaemia; male rat infertility;
KW   hypertension.
XX
OS   Homo sapiens.
XX
PN   WO9940189-A2.
XX
PD   12-AUG-1999.
XX
PF   09-FEB-1999;   99WO-IB00282.
XX
PR   09-FEB-1998;   98US-0074121.
PR   13-APR-1998;   98US-0081563.
PR   10-AUG-1998;   98US-0096116.
PR   04-SEP-1998;   98US-0099273.
XX
PA   (GEST ) GENSET.
XX
PI   Bougueleret L,  Duclert A,  Dumas Milne Edwards J;
XX
DR   WPI; 1999-600966/51.
DR   N-PSDB; AAZ40846.
XX
PT   Extended cDNAs useful for expressing secreted proteins and to obtain
PT   specific antibodies -
XX
PS   Claim 10; Page 240; 244pp; English.
XX
CC   This sequence represents a human secreted protein of the invention.
CC   The extended cDNAs (or genomic DNAs obtainable from them) may be used to
CC   prepare PCR primers and probes. These are useful for forensic matching or
CC   positive identification by DNA sequencing. They may also be used in
CC   alternative fingerprint identification techniques. Antibodies against the
CC   proteins encoded by the extended cDNAs are useful in identification of
CC   tissue types or cell species, as well as identifying tissue specific
```

CC soluble proteins. The sequences can be used for chromosome mapping and
CC identification of genes associated with hereditary diseases or drug
CC response. signal sequences from the cDNAs can be used in construction of
CC secretion vectors. Other sequences derived from the extended cDNAs can be
CC used to clone upstream genomic DNA sequences including promoters. This is
CC in turn useful for identifying proteins that interact with promoter
CC sequences. Some of the proteins may be useful in diagnosing and treating
CC several disorders including, but not limited to: cancer, hyperlipidaemia,
CC cardiovascular and neurodegenerative disorders, autoimmune diseases, and
CC rheumatic diseases, embryogenic disorders, hypertension, renal injury,
CC amino acidurias, hypoglycaemia, male rat infertility and myopathies.
XX
SQ Sequence 131 AA;

Query Match 28.6%; Score 59; DB 20; Length 131;
Best Local Similarity 100.0%; Pred. No. 6.9e-50;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 32 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYE 90
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYE 59

RESULT 15
AAY11881
ID AAY11881 standard; Protein; 56 AA.
XX
AC AAY11881;
XX
DT 18-JUN-1999 (first entry)
XX
DE Human 5' EST secreted protein SEQ ID No: 481.
XX
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide; prostate;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition.
XX
OS Homo sapiens.
XX
PN WO9906550-A2.
XX
PD 11-FEB-1999.
XX
PF 31-JUL-1998; 98WO-IB01232.
XX
PR 01-AUG-1997; 97US-0905144.
XX
PA (GEST) GENSET.
XX
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
XX
DR WPI; 1999-153780/13.
DR N-PSDB; AAX40603.
XX
PT New isolated prostate-derived nucleic acids - used to develop
PT products which may have cytokine, immune regulatory, haematopoiesis
PT regulating, anti-inflammatory or tumour inhibition activity
XX
PS Claim 34; Page 603; 675pp; English.
XX
CC AAX40438 to AAX40715 represent 5' expressed sequence tags (ESTs) for
CC human secreted proteins expressed in prostate, and encode the proteins
CC given in AAY11716 to AAY11993 respectively. The proteins given represent
CC the signal peptide and an N-terminal fragment of a secreted protein. The
CC nucleic acid sequences can be used for producing secreted human gene
CC products. They can also be used to develop products for diagnosis and
CC therapy. The proteins obtained may have cytokine activity, cell
CC proliferation and differentiation activity, haematopoiesis regulating
CC activity, tissue growth regulating activity, reproductive hormone
CC regulating activity, chemotactic/chemokinetic activity, haemostatic and
CC thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, tumour inhibition activity or other activities. The products
CC can be used in forensic, gene therapy and chromosome mapping procedures.
CC The sequences can also be used for obtaining corresponding promoter
CC sequences. The nucleic acids encoding the signal peptides can be used for
CC directing extracellular secretion of a polypeptide or the insertion of a
CC polypeptide into a membrane, or importing a polypeptide into a cell.
XX
SQ Sequence 56 AA;

Query Match 26.7%; Score 55; DB 20; Length 56;
Best Local Similarity 100.0%; Pred. No. 2.6e-46;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 32 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWT 86
|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWT 55

Search completed: October 9, 2002, 16:42:59
Job time : 32 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 16:41:47 ; Search time 21 Seconds
(without alignments)
942.591 Million cell updates/sec

Title: US-09-674-266A-181
Perfect score: 206
Sequence: 1 RLSCAGTLSGSGPHPSRRLT.....DTALLLDNMKKALKLLKTEL 206

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283138 seqs, 96089334 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	175	85.0	175	2	JE0350	Anterior gradient-
2	8	3.9	315	2	JC5201	chemoreceptor TB56
3	8	3.9	637	2	C75578	probable acyl-CoA
4	7	3.4	101	2	C72486	hypothetical prote
5	7	3.4	112	1	XLHU	colipase precursor
6	7	3.4	112	2	B95024	conserved domain p
7	7	3.4	114	2	D90784	hypothetical prote
8	7	3.4	114	2	E90903	hypothetical prote
9	7	3.4	114	2	E85644	hypothetical prote
10	7	3.4	114	2	E85713	unknown protein en
11	7	3.4	115	2	H83514	conserved hypothet
12	7	3.4	126	2	F90432	hypothetical prote
13	7	3.4	142	2	B41132	collagen-related p
14	7	3.4	167	2	AH0652	conserved hypothet
15	7	3.4	205	2	B97895	hypothetical prote
16	7	3.4	235	1	EWWKEC	attacin precursor,
17	7	3.4	251	2	T10262	mast cell serine p
18	7	3.4	258	2	G64992	hypothetical prote
19	7	3.4	258	2	D91017	hypothetical prote
20	7	3.4	258	2	F85861	hypothetical prote
21	7	3.4	261	2	E69053	hypothetical prote
22	7	3.4	272	2	S24375	hypothetical prote
23	7	3.4	277	2	E84478	hypothetical prote
24	7	3.4	301	2	D81745	probable phosphati
25	7	3.4	318	2	A84466	hypothetical prote
26	7	3.4	322	2	B98193	sugar ABC transpor
27	7	3.4	322	2	AH3093	hypothetical prote
28	7	3.4	324	2	T52609	cysteine synthase
29	7	3.4	327	2	S48416	hypothetical prote
30	7	3.4	352	2	S47006	zinc finger protei
31	7	3.4	369	2	AG0386	5-amino-6-(5-phosp
32	7	3.4	381	1	S15809	cytochrome P450 CY
33	7	3.4	386	2	S53965	hypothetical prote
34	7	3.4	390	2	T22810	hypothetical prote
35	7	3.4	396	2	H82339	conserved hypothet
36	7	3.4	402	2	JC5151	nitric-oxide reduc
37	7	3.4	404	1	JC5150	nitric-oxide reduc
38	7	3.4	408	2	PD0007	cytochrome P450 no
39	7	3.4	408	2	JC5674	cytochrome P450 no
40	7	3.4	449	2	AE1952	hypothetical prote
41	7	3.4	455	2	T34412	hypothetical prote
42	7	3.4	465	2	B85358	SERINE CARBOXYPEPT
43	7	3.4	473	2	S22621	phosphomannomutase
44	7	3.4	478	2	S18158	lipoprotein lipase
45	7	3.4	484	2	T34016	hypothetical prote

ALIGNMENTS

RESULT 1
JE0350
Anterior gradient-2 - human
C;Species: Homo sapiens (man)
C;Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 21-Jul-2000
C;Accession: JE0350
R;Thompson, D.A.; Weigel, R.J. .
Biochem. Biophys. Res. Commun. 251, 111-116, 1998
A;Title: hAG-2, the human homologue of the Xenopus laevis cement gland gene XAG-2, is
A;Reference number: JE0350; MUID:99009231
A;Accession: JE0350
A;Molecule type: mRNA
A;Residues: 1-175 <THO>
A;Cross-references: GB:AF007791; NID:g3779196; PIDN:AAC77358.1; PID:g3779197
C;Comment: This protein is coexpressed with estrogen receptor (ER) in breast cancer c
C;Genetics:
A;Gene: hAG-2
C;Superfamily: human anterior gradient-2

Query Match 85.0%; Score 175; DB 2; Length 175;
Best Local Similarity 100.0%; Pred. No. 9.4e-168;
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      32 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 91
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 60

Qy      92 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 151
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 120

Qy     152 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 206
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 175
```

RESULT 2
JC5201
chemoreceptor TB567 - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 20-Feb-1997 #sequence_revision 27-Feb-1997 #text_change 26-Aug-1999
C;Accession: JC5201; PC4303
R;Thomas, M.B.; Haines, S.L.; Akeson, R.A.
Gene 178, 1-5, 1996
A;Title: Chemoreceptors expressed in taste, olfactory and male reproductive tissues.
A;Reference number: JC5200; MUID:97080538
A;Accession: JC5201
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-315 <THO1>
A;Cross-references: GB:U50948; NID:g1256390; PIDN:AAC52910.1; PID:g1256391
A;Accession: PC4303

```
A;Status: preliminary
A;Molecule type: protein
A;Residues: 144-151;270-277 <THO2>
A;Experimental source: taste bud
C;Comment: This protein is coupled to a GTP-binding protein-mediated signal transduction
C;Genetics:
A;Gene: tb567
C;Superfamily: olfactory receptor OR14
C;Keywords: olfaction; taste bud; transmembrane protein
F;26-49/Domain: transmembrane #status predicted <TM1>
F;58-79/Domain: transmembrane #status predicted <TM2>
F;101-120/Domain: transmembrane #status predicted <TM3>
F;140-164/Domain: transmembrane #status predicted <TM4>
F;197-219/Domain: transmembrane #status predicted <TM5>
F;238-260/Domain: transmembrane #status predicted <TM6>
F;273-293/Domain: transmembrane #status predicted <TM7>

  Query Match             3.9%;  Score 8;  DB 2;  Length 315;
  Best Local Similarity   100.0%;  Pred. No. 6.5;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      41 LLLVALSY 48
           ||||||||
Db     142 LLLVALSY 149


RESULT 3
C75578
probable acyl-CoA dehydrogenase - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C;Accession: C75578
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896
A;Accession: C75578
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-637 <WHI>
A;Cross-references: GB:AE001863; GB:AE001825; NID:g6460670; PIDN:AAF12388.1; PID:g646068
A;Experimental source: strain R1
C;Genetics:
A;Gene: DRA0250
A;Map position: 2

  Query Match             3.9%;  Score 8;  DB 2;  Length 637;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     186 ADTALLLD 193
           ||||||||
Db     422 ADTALLLD 429


RESULT 4
C72486
hypothetical protein APE2531 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jun-2000
C;Accession: C72486
R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
A;Reference number: A72450; MUID:99310339
A;Accession: C72486
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-101 <KAW>
A;Cross-references: DDBJ:AP000064; NID:g5105945; PIDN:BAA81547.1; PID:d1045333; PID:g510
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE2531
C;Superfamily: Aeropyrum pernix hypothetical protein APE2531

  Query Match             3.4%;  Score 7;  DB 2;  Length 101;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 AGTLSGS 11
             |||||||
Db      11 AGTLSGS 17


RESULT 5
XLHU
colipase precursor [validated] - human
N;Alternate names: procolipase
C;Species: Homo sapiens (man)
C;Date: 04-Dec-1986 #sequence_revision 19-May-1995 #text_change 08-Dec-2000
C;Accession: A42568; A33949; A03163
R;Sims, H.F.; Lowe, M.E.
Biochemistry 31, 7120-7125, 1992
A;Title: The human colipase gene: isolation, chromosomal location, and tissue-specifi
A;Reference number: A42568; MUID:92353041
A;Accession: A42568
A;Molecule type: DNA
A;Residues: 1-112 <SIM>
A;Cross-references: GB:M95529; NID:g180842; PIDN:AAB05818.1; PID:g1483624
A;Note: sequence extracted from NCBI backbone (NCBIN:110576, NCBIN:110578, NCBIP:1105
R;Lowe, M.E.; Rosenblum, J.L.; McEwen, P.; Strauss, A.W.
Biochemistry 29, 823-828, 1990
A;Title: Cloning and characterization of the human colipase cDNA.
A;Reference number: A33949; MUID:90248429
A;Accession: A33949
A;Molecule type: mRNA
A;Residues: 1-112 <LOW>
A;Cross-references: GB:J02883; NID:g180885; PIDN:AAA52054.1; PID:g180886
A;Note: evidence of partial N-glycosylation, possibly at Asn-43
R;Sternby, B.; Engstrom, A.; Hellman, U.; Vihert, A.M.; Sternby, N.H.; Borgstrom, B.
Biochim. Biophys. Acta 784, 75-80, 1984
A;Title: The primary sequence of human pancreatic colipase.
A;Reference number: A90652; MUID:84104937
A;Accession: A03163
A;Molecule type: protein
A;Residues: 23-108 <STE>
C;Comment: Colipase, a cofactor of triacylglycerol lipase (EC 3.1.1.3), forms a 1:1 s
se the enzyme is washed off by bile salts, which are known to have an inhibitory effe
C;Genetics:
A;Gene: GDB:CLPS
A;Cross-references: GDB:127277; OMIM:120105
A;Map position: 6pter-6p21.1
A;Introns: 28/3; 69/3
C;Superfamily: colipase
C;Keywords: lipid digestion; lipid hydrolysis; pancreas
F;1-17/Domain: signal sequence #status predicted <SIG>
F;18-22/Domain: amino-terminal propeptide #status predicted <APP>
F;23-108/Product: colipase #status experimental <MAT>
F;109-112/Domain: carboxyl-terminal propeptide #status predicted <CPP>
F;34-104,40-56,44-80,45-78,66-86/Disulfide bonds: #status predicted
F;69,72,75,76/Binding site: micellar substrate (Lys, Tyr, Tyr, Tyr) #status predicted

  Query Match             3.4%;  Score 7;  DB 1;  Length 112;
  Best Local Similarity   100.0%;  Pred. No. 28;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      41 LLLVALS 47
           |||||||
Db       7 LLLVALS 13


RESULT 6
B95024
```

conserved domain protein SP0207 [imported] - Streptococcus pneumoniae (strain TIGR4)
C;Species: Streptococcus pneumoniae
C;Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C;Accession: B95024
R;Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A;Reference number: A95000; MUID:21357209; PMID:11463916
A;Accession: B95024
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-112 <KUR>
A;Cross-references: GB:AE005672; PIDN:AAK74387.1; PID:g14971675; GSPDB:GN00164; TIGR:SP4
A;Experimental source: strain TIGR4
C;Genetics:
A;Gene: SP0207

  Query Match 3.4%; Score 7; DB 2; Length 112;
  Best Local Similarity 100.0%; Pred. No. 28;
  Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 50 LARDTTV 56
       |||||||
Db 60 LARDTTV 66

RESULT 7
D90784
hypothetical protein ECs1244 [imported] - Escherichia coli (strain O157:H7, substrain RI
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C;Accession: D90784
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: D90784
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-114 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB34667.1; PID:g13360704; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs1244
C;Superfamily: hypothetical protein b3024

  Query Match 3.4%; Score 7; DB 2; Length 114;
  Best Local Similarity 100.0%; Pred. No. 28;
  Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 197 KALKLLK 203
       |||||||
Db 108 KALKLLK 114

RESULT 8
E90903
hypothetical protein ECs2197 [imported] - Escherichia coli (strain O157:H7, substrain RI
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C;Accession: E90903
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: E90903
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-114 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB35620.1; PID:g13361663; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs2197
C;Superfamily: hypothetical protein b3024

  Query Match 3.4%; Score 7; DB 2; Length 114;
  Best Local Similarity 100.0%; Pred. No. 28;
  Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 197 KALKLLK 203
       |||||||
Db 108 KALKLLK 114

RESULT 9
E85644
hypothetical protein Z1498 [imported] - Escherichia coli (strain O157:H7, substrain E
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: E85644
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: E85644
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-114 <STO>
A;Cross-references: GB:AE005174; NID:g12514358; PIDN:AAG55617.1; GSPDB:GN00145; UWGP:
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z1498
C;Superfamily: hypothetical protein b3024

  Query Match 3.4%; Score 7; DB 2; Length 114;
  Best Local Similarity 100.0%; Pred. No. 28;
  Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 197 KALKLLK 203
       |||||||
Db 108 KALKLLK 114

RESULT 10
E85713
unknown protein encoded within prophage CP-933O [imported] - Escherichia coli (strain
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: E85713
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: E85713
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-114 <STO>
A;Cross-references: GB:AE005174; NID:g12515051; PIDN:AAG56169.1; GSPDB:GN00145; UWGP:
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z2099
C;Superfamily: hypothetical protein b3024

  Query Match 3.4%; Score 7; DB 2; Length 114;
  Best Local Similarity 100.0%; Pred. No. 28;
  Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 197 KALKLLK 203
       |||||||

```
Db    108 KALKLLK 114


RESULT 11
H83514
conserved hypothetical protein PA1059 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 15-Feb-2002
C;Accession: H83514
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathc
A;Reference number: A82950; MUID:20437337
A;Accession: H83514
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-115 <STO>
A;Cross-references: GB:AE004537; GB:AE004091; NID:g9946960; PIDN:AAG04448.1; GSPDB:GN001
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA1059
C;Superfamily: Pyrococcus horikoshii hypothetical protein PH1428

  Query Match             3.4%;  Score 7;  DB 2;  Length 115;
  Best Local Similarity   100.0%;  Pred. No. 28;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      37 VSAFLLL 43
           |||||||
Db      14 VSAFLLL 20


RESULT 12
F90432
hypothetical protein SSO2594 [imported] - Sulfolobus solfataricus
C;Species: Sulfolobus solfataricus
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C;Accession: F90432
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
 Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A;Description: Sulfolobus solfataricus complete genome.
A;Reference number: A99139
A;Accession: F90432
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-126 <KUR>
A;Cross-references: GB:AE006641; NID:g13815896; PIDN:AAK42717.1; GSPDB:GN00155
C;Genetics:
A;Gene: SSO2594

  Query Match             3.4%;  Score 7;  DB 2;  Length 126;
  Best Local Similarity   100.0%;  Pred. No. 31;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      37 VSAFLLL 43
           |||||||
Db      85 VSAFLLL 91


RESULT 13
B41132
collagen-related protein 2 - Hydra magnipapillata (fragment)
C;Species: Hydra magnipapillata
C;Date: 05-Jun-1992 #sequence_revision 05-Jun-1992 #text_change 01-Dec-2000
C;Accession: B41132; S21930
R;Kurz, E.M.; Holstein, T.W.; Petri, B.M.; Engel, J.; David, C.N.
J. Cell Biol. 115, 1159-1169, 1991
A;Title: Mini-collagens in hydra nematocytes.
A;Reference number: A41132; MUID:92064646
A;Accession: B41132
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-142 <KUR>
A;Cross-references: EMBL:X61046; NID:g9448; PIDN:CAA43380.1; PID:g9449
A;Note: submitted to the EMBL Data Library, July 1991
C;Superfamily: unassigned collagens

  Query Match             3.4%;  Score 7;  DB 2;  Length 142;
  Best Local Similarity   100.0%;  Pred. No. 34;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      39 AFLLLVA 45
           |||||||
Db       2 AFLLLVA 8


RESULT 14
AH0652
conserved hypothetical protein STY1322 [imported] - Salmonella enterica subsp. enteri
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001
C;Accession: AH0652
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se
A;Reference number: AB0502; PMID:11677608
A;Accession: AH0652
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-167 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD08403.1; PID:g16502446; GSPDB:GN00176
C;Genetics:
A;Gene: STY1322

  Query Match             3.4%;  Score 7;  DB 2;  Length 167;
  Best Local Similarity   100.0%;  Pred. No. 39;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     197 KALKLLK 203
           |||||||
Db     131 KALKLLK 137


RESULT 15
B97895
hypothetical protein spr0186 [imported] - Streptococcus pneumoniae (strain R6)
C;Species: Streptococcus pneumoniae
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C;Accession: B97895
R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczak, L.; Burgett, S.; DeHoff, B.S.
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.
A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A;Reference number: A97872; MUID:21429245; PMID:11544234
A;Accession: B97895
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-205 <KUR>
A;Cross-references: GB:AE007317; PIDN:AAK98990.1; PID:g15457730; GSPDB:GN00174
C;Genetics:
A;Gene: spr0186

  Query Match             3.4%;  Score 7;  DB 2;  Length 205;
  Best Local Similarity   100.0%;  Pred. No. 46;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      50 LARDTTV 56
           |||||||
Db     153 LARDTTV 159
```

Search completed: October 9, 2002, 16:44:40
Job time : 22 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 16:40:37 ; Search time 14 Seconds
(without alignments)
569.731 Million cell updates/sec

Title: US-09-674-266A-181
Perfect score: 206
Sequence: 1 RLSCAGTLSGSGPHPSRRLT.....DTALLLDNMKKALKLLKTEL 206

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 105224 seqs, 38719550 residues

Word size : 0

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	7	3.4	107	1	COL_RABIT	P42890 oryctolagus
2	7	3.4	112	1	COL_HUMAN	P04118 homo sapien
3	7	3.4	183	1	XAG_XENLA	P55868 xenopus lae
4	7	3.4	185	1	NP77_XENLA	P55869 xenopus lae
5	7	3.4	235	1	ATTE_HYACE	P01513 hyalophora
6	7	3.4	245	1	MCT1_SHEEP	P80931 ovis aries
7	7	3.4	251	1	MCT3_SHEEP	O46683 ovis aries
8	7	3.4	258	1	YFAP_ECOLI	P76462 escherichia
9	7	3.4	302	1	S1C2_HUMAN	O75897 homo sapien
10	7	3.4	327	1	YRB2_YEAST	P40517 saccharomyc
11	7	3.4	352	1	GCS1_YEAST	P35197 saccharomyc
12	7	3.4	381	1	CPXG_STRSQ	P23296 streptomyce
13	7	3.4	386	1	YMK7_YEAST	Q03760 saccharomyc
14	7	3.4	402	1	NOR_FUSOX	P23295 fusarium ox
15	7	3.4	403	1	NOR1_CYLTO	Q00616 cylindrocar
16	7	3.4	408	1	NOR2_CYLTO	Q12599 cylindrocar
17	7	3.4	473	1	RFBM_SALMU	Q00473 salmonella
18	7	3.4	478	1	LIPL_PIG	P49923 sus scrofa
19	7	3.4	514	1	VMT2_HUMAN	Q05940 homo sapien
20	7	3.4	515	1	VMT2_RAT	Q01827 rattus norv
21	7	3.4	517	1	VMT2_BOVIN	Q27963 bos taurus
22	7	3.4	521	1	VMT1_RAT	Q01818 rattus norv
23	7	3.4	525	1	VMT1_HUMAN	P54219 homo sapien
24	7	3.4	532	1	UN17_CAEEL	P34711 caenorhabdi
25	7	3.4	617	1	OM70_YEAST	P07213 saccharomyc
26	7	3.4	705	1	CWBA_BACSU	Q02113 bacillus su
27	7	3.4	721	1	RIR1_MYCGE	P47473 mycoplasma
28	7	3.4	721	1	RIR1_MYCPN	P78027 mycoplasma
29	7	3.4	786	1	FL10_CHLRE	P46869 chlamydomon
30	7	3.4	787	1	SP3E_BACSU	P21458 bacillus su
31	7	3.4	1443	1	DPO3_MYCPN	P75080 mycoplasma
32	6	2.9	20	1	CPA7_PAPSP	P80055 papio sp. (
33	6	2.9	74	1	WDNM_RAT	P14730 rattus norv
34	6	2.9	88	1	VG85_BPML5	Q05302 mycobacteri
35	6	2.9	89	1	FLIQ_ECOLI	P33134 escherichia
36	6	2.9	89	1	FLIQ_SALTY	P54701 salmonella
37	6	2.9	102	1	CMGC_BACHD	Q9k923 bacillus ha
38	6	2.9	102	1	SP12_HUMAN	Q9y6a9 homo sapien
39	6	2.9	102	1	SP12_MOUSE	Q9d958 mus musculu
40	6	2.9	103	1	HE2_HUMAN	Q08648 homo sapien
41	6	2.9	104	1	GROA_BOVIN	O46676 bos taurus
42	6	2.9	104	1	RL3E_SULAC	P11522 sulfolobus
43	6	2.9	106	1	COLA_HORSE	P02704 equus cabal
44	6	2.9	107	1	IF1C_MAIZE	P46618 zea mays (m
45	6	2.9	107	1	IF1C_ORYSA	P12135 oryza sativ

ALIGNMENTS

RESULT 1
COL_RABIT
ID COL_RABIT STANDARD; PRT; 107 AA.
AC P42890;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Colipase precursor.
GN CLPS.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=93345715; PubMed=8344444;
RA Colwell N.S., Aleman-Gomez J.A., Sasser T.L., Kumar V.B.;
RT "Cloning and characterization of rabbit pancreatic colipase.";
RL Int. J. Biochem. 25:885-890(1993).
CC -!- FUNCTION: COLIPASE IS A COFACTOR OF PANCREATIC LIPASE. IT ALLOWS
CC THE LIPASE TO ANCHOR ITSELF TO THE LIPID-WATER INTERFACE. WITHOUT
CC COLIPASE THE ENZYME IS WASHED OFF BY BILE SALTS, WHICH HAVE AN
CC INHIBITORY EFFECT ON THE LIPASE.
CC -!- FUNCTION: ENTEROSTATIN HAS A BIOLOGICAL ACTIVITY AS A SATIETY
CC SIGNAL (BY SIMILARITY).
CC -!- SUBUNIT: FORM A 1:1 STOICHIOMETRIC COMPLEX WITH PANCREATIC LIPASE.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; L06329; AAA02911.1; -.
DR HSSP; P02703; 1PCN.
DR InterPro; IPR001981; Colipase.
DR Pfam; PF01114; Colipase; 1.
DR Pfam; PF02740; Colipase_C; 1.
DR PRINTS; PR00128; COLIPASE.
DR SMART; SM00023; COLIPASE; 1.
DR PROSITE; PS00121; COLIPASE; 1.
KW Lipid degradation; Digestion; Pancreas; Signal.
FT SIGNAL 1 17 POTENTIAL.
FT PROPEP 18 22 ENTEROSTATIN, ACTIVATION PEPTIDE
FT (POTENTIAL).
FT CHAIN 23 107 COLIPASE.
FT DISULFID 34 45 BY SIMILARITY.
FT DISULFID 40 56 BY SIMILARITY.
FT DISULFID 44 78 BY SIMILARITY.
FT DISULFID 66 86 BY SIMILARITY.
FT DISULFID 80 104 BY SIMILARITY.
SQ SEQUENCE 107 AA; 11271 MW; 825BA1AEB1422390 CRC64;

```
  Query Match             3.4%;  Score 7;  DB 1;  Length 107;
  Best Local Similarity   100.0%;  Pred. No. 7.7;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      41 LLLVALS 47
           |||||||
Db       7 LLLVALS 13


RESULT 2
COL_HUMAN
ID   COL_HUMAN      STANDARD;      PRT;   112 AA.
AC   P04118;
DT   01-NOV-1986 (Rel. 03, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Colipase precursor.
GN   CLPS.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90248429; PubMed=2337598;
RA   Lowe M.E., Rosenblum J.L., McEwen P., Strauss A.W.;
RT   "Cloning and characterization of the human colipase cDNA.";
RL   Biochemistry 29:823-828(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92353041; PubMed=1643046;
RA   Sims H.F., Lowe M.E.;
RT   "The human colipase gene: isolation, chromosomal location, and
RT   tissue-specific expression.";
RL   Biochemistry 31:7120-7125(1992).
RN   [3]
RP   SEQUENCE OF 23-108.
RX   MEDLINE=84104937; PubMed=6691986;
RA   Sternby B., Engstroem A., Hellman U., Vihert A.M., Sternby N.-H.,
RA   Borgstroem B.;
RT   "The primary sequence of human pancreatic colipase.";
RL   Biochim. Biophys. Acta 784:75-80(1984).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
RX   MEDLINE=93241293; PubMed=8479519;
RA   van Tilbeurgh H., Egloff M.-P., Martinez C., Rugani N., Verger R.,
RA   Cambillau C.;
RT   "Interfacial activation of the lipase-procolipase complex by mixed
RT   micelles revealed by X-ray crystallography.";
RL   Nature 362:814-820(1993).
CC   -!- FUNCTION: COLIPASE IS A COFACTOR OF PANCREATIC LIPASE. IT ALLOWS
CC       THE LIPASE TO ANCHOR ITSELF TO THE LIPID-WATER INTERFACE. WITHOUT
CC       COLIPASE THE ENZYME IS WASHED OFF BY BILE SALTS, WHICH HAVE AN
CC       INHIBITORY EFFECT ON THE LIPASE.
CC   -!- FUNCTION: ENTEROSTATIN HAS A BIOLOGICAL ACTIVITY AS A SATIETY
CC       SIGNAL.
CC   -!- SUBUNIT: FORM A 1:1 STOICHIOMETRIC COMPLEX WITH PANCREATIC LIPASE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J02883; AAA52054.1; -.
DR   EMBL; M95529; AAB05818.1; -.
DR   PIR; A03163; XLHU.
DR   PIR; A33949; A33949.
DR   PIR; A42568; A42568.
DR   HSSP; P02703; 1ETH.
DR   MIM; 120105; -.
DR   InterPro; IPR001981; Colipase.
DR   Pfam; PF01114; Colipase; 1.
DR   Pfam; PF02740; Colipase_C; 1.
DR   PRINTS; PR00128; COLIPASE.
DR   SMART; SM00023; COLIPASE; 1.
DR   PROSITE; PS00121; COLIPASE; 1.
KW   Lipid degradation; Digestion; Pancreas; Signal.
FT   SIGNAL        1     17
FT   PROPEP       18     22       ENTEROSTATIN, ACTIVATION PEPTIDE
FT                                (POTENTIAL).
FT   CHAIN        23    112       COLIPASE.
FT   DISULFID     34     45       BY SIMILARITY.
FT   DISULFID     40     56       BY SIMILARITY.
FT   DISULFID     44     78       BY SIMILARITY.
FT   DISULFID     66     86       BY SIMILARITY.
FT   DISULFID     80    104       BY SIMILARITY.
FT   CONFLICT     68     69       MISSING (IN REF. 2).
SQ   SEQUENCE   112 AA;  11954 MW;  772872EBBE7C4DF8 CRC64;

  Query Match             3.4%;  Score 7;  DB 1;  Length 112;
  Best Local Similarity   100.0%;  Pred. No. 8;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      41 LLLVALS 47
           |||||||
Db       7 LLLVALS 13


RESULT 3
XAG_XENLA
ID   XAG_XENLA      STANDARD;      PRT;   183 AA.
AC   P55868;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Putative secreted protein XAG precursor.
GN   XAG.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Cement gland;
RX   MEDLINE=97003224; PubMed=8850563;
RA   Sive H., Bradley L.;
RT   "A sticky problem: the Xenopus cement gland as a paradigm for
RT   anteroposterior patterning.";
RL   Dev. Dyn. 205:265-280(1996).
CC   -!- SUBCELLULAR LOCATION: Secreted (Probable).
CC   -!- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVEL IN CEMENT GLAND.
CC   -!- SIMILARITY: HIGH, TO XENOPUS NP77.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U76752; AAB18819.1; -.
KW   Signal.
FT   SIGNAL        1     18       POTENTIAL.
FT   CHAIN        19    183       PUTATIVE SECRETED PROTEIN XAG.
SQ   SEQUENCE   183 AA;  20475 MW;  95E12B06FCB8DCE8 CRC64;

  Query Match             3.4%;  Score 7;  DB 1;  Length 183;
  Best Local Similarity   100.0%;  Pred. No. 13;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     163 DPSLTVR 169
```

```
          |||||||
Db    140 DPSLTVR 146


RESULT 4
NP77_XENLA
ID   NP77_XENLA     STANDARD;      PRT;   185 AA.
AC   P55869;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Putative secreted protein NP77 precursor.
GN   NP77.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Cement gland;
RA   Aberger F., Schueren C., Lepperdinger G., Richter K., Grunz H.;
RL   Submitted (MAR-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Secreted (Probable).
CC   -!- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVEL IN CEMENT GLAND
CC       (POTENTIAL).
CC   -!- SIMILARITY: HIGH, TO XENOPUS XAG.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U82110; AAB49974.1; -.
KW   Signal.
FT   SIGNAL        1     18       POTENTIAL.
FT   CHAIN        19    185       PUTATIVE SECRETED PROTEIN NP77.
SQ   SEQUENCE   185 AA;  20442 MW;  AE3807C926044509 CRC64;

  Query Match             3.4%;  Score 7;  DB 1;  Length 185;
  Best Local Similarity   100.0%;  Pred. No. 13;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    163 DPSLTVR 169
          |||||||
Db    142 DPSLTVR 148


RESULT 5
ATTE_HYACE
ID   ATTE_HYACE     STANDARD;      PRT;   235 AA.
AC   P01513;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Attacin E and F precursor (Immune protein P5).
OS   Hyalophora cecropia (Cecropia moth).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Bombycoidea; Saturniidae; Saturniinae; Hyalophora.
OX   NCBI_TaxID=7123;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91160561; PubMed=2001705;
RA   Sun S.C., Lindstroem I., Lee J.-Y., Faye I.;
RT   "Structure and expression of the attacin genes in Hyalophora
RT   cecropia.";
RL   Eur. J. Biochem. 196:247-254(1991).
RN   [2]
RP   SEQUENCE OF 48-235 FROM N.A.
RA   Kockum K., Faye I., von Hofsten P., Lee J.-Y., Xanthopoulos K.G.,
RA   Boman H.G.;
RT   "Insect immunity. Isolation and sequence of two cDNA clones
RT   corresponding to acidic and basic attacins from Hyalophora cecropia.";
RL   EMBO J. 3:2071-2075(1984).
RN   [3]
RP   SEQUENCE OF 48-235 FROM N.A.
RX   MEDLINE=86005745; PubMed=3840100;
RA   Boman H.G., Faye I., von Hofsten P., Kockum K., Lee J.-Y.,
RA   Xanthopoulos K.G., Bennich H., Engstroem A., Merrifield R.B.,
RA   Andreu D.;
RT   "On the primary structures of lysozyme, cecropins and attacins from
RT   Hyalophora cecropia.";
RL   Dev. Comp. Immunol. 9:551-558(1985).
RN   [4]
RP   SEQUENCE OF 48-231.
RA   Engstroem A., Engstroem P., Tao Z.-J., Carlsson A., Bennich H.;
RT   "Insect immunity. The primary structure of the antibacterial protein
RT   attacin F and its relation to two native attacins from Hyalophora
RT   cecropia.";
RL   EMBO J. 3:2065-2070(1984).
CC   -!- FUNCTION: ATTACINS ARE HEMOLYMPH ANTIBACTERIAL PROTEINS.
CC   -!- PTM: ATTACIN F APPEARS TO BE DERIVED BY PROTEOLYTIC DIGESTION OF
CC       ATTACIN E.
CC   -!- MISCELLANEOUS: THERE ARE SIX FORMS OF ATTACIN THAT ARE DIVIDED
CC       INTO TWO GROUPS: ACIDIC (E AND F) AND BASIC (A, B, C, AND D).
CC   -!- SIMILARITY: BELONGS TO THE ATTACIN/SARCOTOXIN II FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X57715; CAA40886.1; -.
DR   EMBL; X00869; CAA25414.1; -.
DR   EMBL; M34926; AAA29183.1; -.
DR   PIR; A01775; EWWKEC.
DR   PIR; S14104; S14104.
KW   Insect immunity; Antibiotic; Hemolymph; Multigene family; Signal;
KW   Repeat.
FT   SIGNAL        1     19       POTENTIAL.
FT   PROPEP       20     47
FT   CHAIN        48    235       ATTACIN E.
FT   CHAIN        48    231       ATTACIN F.
FT   DOMAIN      104    169       GLY-RICH (G1).
FT   DOMAIN      170    235       GLY-RICH (G2).
SQ   SEQUENCE   235 AA;  25437 MW;  35E1B549FB0417D1 CRC64;

  Query Match             3.4%;  Score 7;  DB 1;  Length 235;
  Best Local Similarity   100.0%;  Pred. No. 16;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     40 FLLLVAL 46
          |||||||
Db      7 FLLLVAL 13


RESULT 6
MCT1_SHEEP
ID   MCT1_SHEEP     STANDARD;      PRT;   245 AA.
AC   P80931;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Mast cell protease 1A precursor (EC 3.4.21.-) (SMCP-1A).
OS   Ovis aries (Sheep).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Caprinae; Ovis.
OX   NCBI_TaxID=9940;
```

```
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 176-185.
RC   TISSUE=Mast cells;
RX   MEDLINE=98343972; PubMed=9677343;
RA   McAleese S.M., Pemberton A.D., McGrath M.E., Huntley J.F.,
RA   Miller H.R.P.;
RT   "Sheep mast-cell proteinases-1 and -3: cDNA cloning, primary
RT   structure and molecular modelling of the enzymes and further studies
RT   on substrate specificity.";
RL   Biochem. J. 333:801-809(1998).
RN   [2]
RP   SEQUENCE OF 20-44.
RA   Miller H.R.P., Huntley J.F., Newlands G.F.J.;
RL   (In) Caughey G.H. (eds.);
RL   Mast cell proteases in immunology and biology, pp.203-235, Marcel
RL   Dekker, New York (1995).
RN   [3]
RP   SEQUENCE OF 20-29.
RC   TISSUE=Gastric mucosa;
RX   MEDLINE=97184650; PubMed=9032451;
RA   Pemberton A.D., Huntley J.F., Miller H.R.P.;
RT   "Sheep mast cell proteinase-1: characterization as a member of a new
RT   class of dual-specific ruminant chymases.";
RL   Biochem. J. 321:665-670(1997).
CC   -!- FUNCTION: HAS A CHYMOTRYPSIN-LIKE AND TRYPSIN-LIKE ACTIVITY.
CC   -!- SUBCELLULAR LOCATION: Secretory granules.
CC   -!- TISSUE SPECIFICITY: MUCOSAL MAST CELLS.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY. GRANZYME SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y14654; CAA74984.1; -.
DR   HSSP; P04187; 2CP1.
DR   MEROPS; S01.142; -.
DR   InterPro; IPR001314; Chymotrypsin.
DR   InterPro; IPR001254; Trypsin.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
KW   Hydrolase; Serine protease; Zymogen; Signal.
FT   SIGNAL       1     17       POTENTIAL.
FT   PROPEP      18     19       ACTIVATION PEPTIDE.
FT   CHAIN       20    245       MAST CELL PROTEASE 1A.
FT   ACT_SITE    63     63       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   107    107       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   201    201       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   DISULFID    48     64       BY SIMILARITY.
FT   DISULFID   141    207       BY SIMILARITY.
FT   DISULFID   172    186       BY SIMILARITY.
FT   CONFLICT    20     20       I -> F (IN REF. 2; AA SEQUENCE).
SQ   SEQUENCE   245 AA;  26877 MW;  C362CF3367FFF180 CRC64;

  Query Match             3.4%;  Score 7;  DB 1;  Length 245;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      40 FLLLVAL 46
           |||||||
Db       4 FLLLVAL 10


RESULT 7
MCT3_SHEEP
ID   MCT3_SHEEP     STANDARD;      PRT;   251 AA.
AC   O46683;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Mast cell protease 3 precursor (EC 3.4.21.-) (SMCP-3).
OS   Ovis aries (Sheep).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Caprinae; Ovis.
OX   NCBI_TaxID=9940;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Bone marrow;
RX   MEDLINE=98343972; PubMed=9677343;
RA   McAleese S.M., Pemberton A.D., McGrath M.E., Huntley J.F.,
RA   Miller H.R.P.;
RT   "Sheep mast-cell proteinases-1 and -3: cDNA cloning, primary
RT   structure and molecular modelling of the enzymes and further studies
RT   on substrate specificity.";
RL   Biochem. J. 333:801-809(1998).
CC   -!- FUNCTION: HAS A CHYMOTRYPSIN-LIKE AND TRYPSIN-LIKE ACTIVITY.
CC   -!- SUBCELLULAR LOCATION: Secretory granules.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY. GRANZYME SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y13462; CAA73859.1; -.
DR   HSSP; P00763; 1DPO.
DR   InterPro; IPR001314; Chymotrypsin.
DR   InterPro; IPR001254; Trypsin.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
KW   Hydrolase; Serine protease; Zymogen; Signal.
FT   SIGNAL       1     17       POTENTIAL.
FT   PROPEP      18     19       ACTIVATION PEPTIDE.
FT   CHAIN       20    251       MAST CELL PROTEASE 3.
FT   ACT_SITE    63     63       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   107    107       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   201    201       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   DISULFID    48     64       BY SIMILARITY.
FT   DISULFID   141    207       BY SIMILARITY.
FT   DISULFID   172    186       BY SIMILARITY.
FT   CARBOHYD    70     70       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   251 AA;  27510 MW;  DF2F986114AAE909 CRC64;

  Query Match             3.4%;  Score 7;  DB 1;  Length 251;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      40 FLLLVAL 46
           |||||||
Db       4 FLLLVAL 10


RESULT 8
YFAP_ECOLI
ID   YFAP_ECOLI     STANDARD;      PRT;   258 AA.
AC   P76462;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
```

```
DE   Hypothetical protein yfaP precursor.
GN   YFAP OR B2225.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000312; AAC75285.1; -.
DR   EcoGene; EG14078; yfaP.
KW   Hypothetical protein; Signal; Complete proteome.
FT   SIGNAL       1     19       POTENTIAL.
FT   CHAIN       20    258       HYPOTHETICAL PROTEIN YFAP.
SQ   SEQUENCE   258 AA;  28303 MW;  911F2B5C433E5F41 CRC64;

  Query Match             3.4%;  Score 7;  DB 1;  Length 258;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      41 LLLVALS 47
           |||||||
Db       8 LLLVALS 14


RESULT 9
S1C2_HUMAN
ID   S1C2_HUMAN     STANDARD;      PRT;   302 AA.
AC   O75897;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Sulfotransferase 1C2 (EC 2.8.2.-) (SULT1C) (SULT1C#2).
GN   SULT1C2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Fetal lung;
RX   MEDLINE=99069375; PubMed=9852044;
RA   Sakakibara Y., Yanagisawa K., Katafuchi J., Ringer D.P., Takami Y.,
RA   Nakayama T., Suiko M., Liu M.-C.;
RT   "Molecular cloning, expression, and characterization of novel human
RT   SULT1C sulfotransferases that catalyze the sulfonation of
RT   N-hydroxy-2-acetylaminofluorene.";
RL   J. Biol. Chem. 273:33929-33935(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20247255; PubMed=10783263;
RA   Freimuth R.R., Raftogianis R.B., Wood T.C., Moon E., Kim U.-J., Xu J.,
RA   Siciliano M.J., Weinshilboum R.M.;
RT   "Human sulfotransferases SULT1C1 and SULT1C2: cDNA characterization,
RT   gene cloning, and chromosomal localization.";
RL   Genomics 65:157-165(2000).
CC   -!- FUNCTION: CATALYZES THE
CC       SULFATE CONJUGATION OF MANY DRUGS, XENOBIOTIC COMPOUNDS, HORMONES,
CC       AND NEUROTRANSMITTERS. MAY BE INVOLVED IN THE ACTIVATION OF
CC       CARCINOGENIC HYROXYLAMINES. SHOWS ACTIVITY TOWARDS P-NITROPHENOL
CC       AND N-HYDROXY-2-ACETYLAMINO-FLUORENE (N-OH-2AAF).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVELS IN FETAL LUNG AND
CC       KIDNEY AND AT LOW LEVELS IN FETAL HEART, ADULT KIDNEY, OVARY AND
CC       SPINAL CHORD.
CC   -!- SIMILARITY: BELONGS TO THE SULFOTRANSFERASES FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF055584; AAC95519.1; -.
DR   EMBL; AF186263; AAF72810.1; -.
DR   HSSP; P50224; 1CJM.
DR   InterPro; IPR000863; Sulfotransferase.
DR   Pfam; PF00685; Sulfotransfer; 1.
DR   ProDom; PD001218; Sulfotransfer; 1.
KW   Transferase.
FT   CONFLICT      5      5       E -> D (IN REF. 2).
SQ   SEQUENCE   302 AA;  35534 MW;  DD2F4865B2E3B56D CRC64;

  Query Match             3.4%;  Score 7;  DB 1;  Length 302;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     127 EIQKLAE 133
           |||||||
Db     210 EIQKLAE 216


RESULT 10
YRB2_YEAST
ID   YRB2_YEAST     STANDARD;      PRT;   327 AA.
AC   P40517;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Ran-specific GTPase-activating protein 2 (Ran binding protein 2)
DE   (RANBP2).
GN   YRB2 OR YIL063C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA   Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA   Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA   Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA   Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA   Walsh S.V., Whitehead S.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE=98058988; PubMed=9395535;
RA   Taura T., Schlenstedt G., Silver P.A.;
RT   "Yrb2p is a nuclear protein that interacts with Prp20p, a yeast Rcc1
RT   homologue.";
RL   J. Biol. Chem. 272:31877-31884(1997).
RN   [3]
RP   CHARACTERIZATION.
RX   MEDLINE=98301591; PubMed=9636166;
RA   Taura T., Krebber H., Silver P.A.;
RT   "A member of the Ran-binding protein family, Yrb2p, is involved in
```

```
RT   nuclear protein export.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:7427-7432(1998).
CC   -!- FUNCTION: Important for the export of protein conaining nuclear
CC       export signal (NES) out of the nucleus. Stimulates the GTPase
CC       activity of GSP1.
CC   -!- SUBUNIT: Interacts with GSP1, XPO1 and PRP20.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DOMAIN: CONTAINS X-F-X-F-G REPEATS.
CC   -!- SIMILARITY: CONTAINS 1 RANBP1-LIKE DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z38060; CAA86160.1; -.
DR   PIR; S48416; S48416.
DR   HSSP; P49792; 1RRP.
DR   SGD; S0001325; YRB2.
DR   InterPro; IPR000697; RanBP1_WASP.
DR   InterPro; IPR000156; Ran_BP1.
DR   Pfam; PF00638; Ran_BP1; 1.
DR   SMART; SM00160; RanBD; 1.
KW   GTPase activation; Nuclear protein; Protein transport; Transport.
SQ   SEQUENCE   327 AA;  36054 MW;  7A1A626B9171646D CRC64;

  Query Match             3.4%;  Score 7;  DB 1;  Length 327;
  Best Local Similarity   100.0%;  Pred. No. 22;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      60 AKKDTKD 66
           |||||||
Db     113 AKKDTKD 119


RESULT 11
GCS1_YEAST
ID   GCS1_YEAST     STANDARD;      PRT;   352 AA.
AC   P35197;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Zinc finger protein GCS1.
GN   GCS1 OR YDL226C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94349929; PubMed=8070409;
RA   Ireland L.S., Johnston G.C., Drebot M.A., Dhillon N., Demaggio A.J.,
RA   Hoekstra M.F., Singer R.A.;
RT   "A member of a novel family of yeast 'zn-finger' proteins mediates
RT   the transition from stationary phase to cell proliferation.";
RL   EMBO J. 13:3812-3821(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Rasmussen S.W.;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: PLAYS A ROLE IN THE RESUMPTION OF YEAST CELL
CC       PROLIFERATION FROM STATIONARY PHASE.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- SIMILARITY: BELONGS TO THE GCS1/GLO3/SPS18 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L24125; AAA50389.1; -.
DR   EMBL; Z74274; CAA98805.1; -.
DR   SGD; S0002385; GCS1.
DR   InterPro; IPR001164; Znf_GCS.
DR   Pfam; PF01412; ArfGap; 1.
DR   PRINTS; PR00405; REVINTRACTNG.
DR   SMART; SM00105; ArfGap; 1.
KW   Zinc-finger; Nuclear protein; DNA-binding.
FT   ZN_FING      26     49       C4-TYPE.
FT   VARIANT      27     27       M -> I.
FT   VARIANT      29     29       C -> Y (IN GCS1-1; SEVERELY AFFECT THE
FT                                FUNCTION).
SQ   SEQUENCE   352 AA;  39296 MW;  3215525F3EF9CF17 CRC64;

  Query Match             3.4%;  Score 7;  DB 1;  Length 352;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      74 TLSRGWG 80
           |||||||
Db     245 TLSRGWG 251


RESULT 12
CPXG_STRSQ
ID   CPXG_STRSQ     STANDARD;      PRT;   381 AA.
AC   P23296;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Cytochrome P450 105C1 (EC 1.14.-.-).
GN   CYP105C1 OR CHOP.
OS   Streptomyces sp.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1931;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90299781; PubMed=2361941;
RA   Horii M., Ishizaki T., Paik S.Y., Manome T., Murooka Y.;
RT   "An operon containing the genes for cholesterol oxidase and a
RT   cytochrome P-450-like protein from a Streptomyces sp.";
RL   J. Bacteriol. 172:3644-3653(1990).
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M31939; AAA26718.1; -.
DR   HSSP; Q00441; 1OXA.
DR   InterPro; IPR001128; Cyt_P450.
DR   Pfam; PF00067; p450; 1.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme.
FT   BINDING     330    330       HEME (BY SIMILARITY).
SQ   SEQUENCE   381 AA;  41703 MW;  841B959C9DDEA99C CRC64;

  Query Match             3.4%;  Score 7;  DB 1;  Length 381;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      39 AFLLLVA 45
           |||||||
Db     213 AFLLLVA 219
```

```
RESULT 13
YMK7_YEAST
ID   YMK7_YEAST     STANDARD;      PRT;   386 AA.
AC   Q03760;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Hypothetical 45.2 kDa protein in ZDS2-URA5 intergenic region.
GN   YML107C OR YM8339.12.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Skelton J., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL   Submitted (MAY-1995) to the EMBL/GenBank/DDBJ databases.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z49210; CAA89111.1; -.
DR   SGD; S0004575; YML107C.
KW   Hypothetical protein.
SQ   SEQUENCE   386 AA;  45182 MW;  0AC670A62AB2EB56 CRC64;

  Query Match             3.4%;  Score 7;  DB 1;  Length 386;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     126 KEIQKLA 132
           |||||||
Db     285 KEIQKLA 291


RESULT 14
NOR_FUSOX
ID   NOR_FUSOX      STANDARD;      PRT;   402 AA.
AC   P23295;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Cytochrome P450 55A1 (EC 1.14.-.-) (CYPLVA1) (P450 DNIR) (Nitric-oxide
DE   reductase) (P450 NOR).
GN   CYP55A1 OR CYP55.
OS   Fusarium oxysporum.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Hypocreales; mitosporic Hypocreales; Fusarium.
OX   NCBI_TaxID=5507;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=MT-811;
RX   MEDLINE=91244845; PubMed=2037602;
RA   Kizawa H., Tomura D., Oda M., Fukamizu A., Hoshino T., Gotoh O.,
RA   Yasui T., Shoun H.;
RT   "Nucleotide sequence of the unique nitrate/nitrite-inducible
RT   cytochrome P-450 cDNA from Fusarium oxysporum.";
RL   J. Biol. Chem. 266:10632-10637(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MT-811;
RX   MEDLINE=95096031; PubMed=7798191;
RA   Tomura D., Obika K., Fukamizu A., Shoun H.;
RT   "Nitric oxide reductase cytochrome P-450 gene, CYP 55, of the fungus
RT   Fusarium oxysporum containing a potential binding-site for FNR, the
RT   transcription factor involved in the regulation of anaerobic growth
RT   of Escherichia coli.";
RL   J. Biochem. 116:88-94(1994).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX   MEDLINE=97475224; PubMed=9334748;
RA   Park S.-Y., Shimizu H., Adachi S.-I., Nakagawa A., Tanaka I.,
RA   Nakahara K., Shoun H., Obayashi E., Nakamura H., Iizuka T., Shiro Y.;
RT   "Crystal structure of nitric oxide reductase from denitrifying fungus
RT   Fusarium oxysporum.";
RL   Nat. Struct. Biol. 4:827-832(1997).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS).
RX   MEDLINE=20138220; PubMed=10671516;
RA   Shimizu H., Obayashi E., Gomi Y., Arakawa H., Park S.-Y., Nakamura H.,
RA   Adachi S.-I., Shoun H., Shiro Y.;
RT   "Proton delivery in NO reduction by fungal nitric-oxide reductase.
RT   Cryogenic crystallography, spectroscopy, and kinetics of ferric-NO
RT   complexes of wild-type and mutant enzymes.";
RL   J. Biol. Chem. 275:4816-4826(2000).
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).
RX   MEDLINE=21159060; PubMed=11258878;
RA   Lee D.-S., Park S.-Y., Yamane K., Obayashi E., Hori H., Shiro Y.;
RT   "Structural characterization of n-butyl-isocyanide complexes of
RT   cytochromes P450nor and P450cam.";
RL   Biochemistry 40:2669-2677(2001).
CC   -!- FUNCTION: INVOLVED IN A DISSIMILATORY REDUCTION OF NITRITE, ACTS
CC       AS A NITRIC OXIDE REDUCTASE.
CC   -!- INDUCTION: BY NITRATE/NITRITE.
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M63340; AAA33337.1; -.
DR   EMBL; D14517; BAA03390.1; -.
DR   PIR; A40401; A40401.
DR   PIR; JX0335; JX0335.
DR   PDB; 1ROM; 15-OCT-97.
DR   PDB; 2ROM; 15-OCT-97.
DR   PDB; 1CL6; 22-MAR-00.
DR   PDB; 1CMJ; 22-MAR-00.
DR   PDB; 1CMN; 22-MAR-00.
DR   PDB; 1F24; 23-NOV-00.
DR   PDB; 1F25; 23-NOV-00.
DR   PDB; 1F26; 23-NOV-00.
DR   PDB; 1GED; 22-NOV-00.
DR   PDB; 1GEI; 29-NOV-00.
DR   PDB; 1GEJ; 06-DEC-00.
DR   InterPro; IPR001128; Cyt_P450.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR00359; BP450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Heme; Electron transport; 3D-structure.
FT   INIT_MET      0      0
FT   BINDING     351    351       HEME.
SQ   SEQUENCE   402 AA;  44240 MW;  6422AFB24FC7213D CRC64;

  Query Match             3.4%;  Score 7;  DB 1;  Length 402;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      39 AFLLLVA 45
           |||||||
Db     232 AFLLLVA 238


RESULT 15
```

```
NOR1_CYLTO
ID   NOR1_CYLTO     STANDARD;      PRT;   403 AA.
AC   Q00616;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Cytochrome P450 55A2 (EC 1.14.-.-) (Cytochrome P450NOR1).
GN   CYP55A2.
OS   Cylindrocarpon tonkinense.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Hypocreales; Nectriaceae; Neonectria.
OX   NCBI_TaxID=42744;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IFO 30561;
RX   MEDLINE=97163854; PubMed=9010609;
RA   Kudo T., Tomura D., Liu D.L., Dai X.Q., Shoun H.;
RT   "Two isozymes of P450nor of Cylindrocarpon tonkinense: molecular
RT   cloning of the cDNAs and genes, expressions in the yeast, and the
RT   putative NAD(P)H-binding site.";
RL   Biochimie 78:792-799(1996).
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; D78511; BAA11408.1; -.
DR   HSSP; P23295; 1F24.
DR   InterPro; IPR001128; Cyt_P450.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR00359; BP450.
DR   PROSITE; PS00086; CYTOCHROME_P450; FALSE_NEG.
KW   Oxidoreductase; Monooxygenase; Heme.
FT   BINDING     352    352       HEME (BY SIMILARITY).
SQ   SEQUENCE   403 AA;  44273 MW;  1C2AB643FD219534 CRC64;

  Query Match             3.4%;  Score 7;  DB 1;  Length 403;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      39 AFLLLVA 45
           |||||||
Db     233 AFLLLVA 239


Search completed: October  9, 2002, 16:43:27
Job time : 15 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 16:41:17 ; Search time 28 Seconds
(without alignments)
1272.748 Million cell updates/sec

Title: US-09-674-266A-181
Perfect score: 206
Sequence: 1 RLSCAGTLSGSGPHPSRRLT.....DTALLLDNMKKALKLLKTEL 206

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 562222 seqs, 172994929 residues

Word size : 0

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	175	85.0	175	4	O95994	O95994 homo sapien
2	34	16.5	175	11	O88312	O88312 mus musculu
3	10	4.9	159	13	Q90Y05	Q90y05 xenopus lae
4	8	3.9	200	16	Q988W1	Q988w1 rhizobium l
5	8	3.9	315	11	Q62943	Q62943 rattus norv
6	8	3.9	637	16	Q9RYQ8	Q9ryq8 deinococcus
7	8	3.9	1232	5	Q960R8	Q960r8 drosophila
8	8	3.9	1284	5	Q9VB20	Q9vb20 drosophila
9	7	3.4	73	12	Q91ER9	Q91er9 cydia pomon
10	7	3.4	80	8	O47957	O47957 phoxinus eo
11	7	3.4	101	2	Q56157	Q56157 streptomyce
12	7	3.4	101	17	Q9Y8V3	Q9y8v3 aeropyrum p
13	7	3.4	109	4	Q9GZT3	Q9gzt3 homo sapien
14	7	3.4	112	16	Q97SV7	Q97sv7 streptococc
15	7	3.4	114	2	Q9KXA4	Q9kxa4 escherichia
16	7	3.4	114	9	Q9XJM3	Q9xjm3 bacteriopha
17	7	3.4	115	16	Q9I4R5	Q9i4r5 pseudomonas
18	7	3.4	120	3	Q05715	Q05715 saccharomyc
19	7	3.4	126	17	Q97VM7	Q97vm7 sulfolobus
20	7	3.4	142	5	Q00485	Q00485 hydra sp. m
21	7	3.4	167	2	Q9L435	Q9l435 salmonella
22	7	3.4	167	5	Q9V6R3	Q9v6r3 drosophila
23	7	3.4	169	17	Q96ZA8	Q96za8 sulfolobus
24	7	3.4	173	8	Q951J1	Q951j1 phoxinus er
25	7	3.4	173	8	Q951I9	Q951i9 phoxinus er
26	7	3.4	185	13	O42251	O42251 xenopus lae
27	7	3.4	188	5	Q95XX7	Q95xx7 caenorhabdi
28	7	3.4	227	2	Q9EWR7	Q9ewr7 streptomyce
29	7	3.4	257	10	Q9FJ08	Q9fj08 arabidopsis
30	7	3.4	257	12	Q9J2L8	Q9j2l8 macaca mula
31	7	3.4	261	17	O27451	O27451 methanother
32	7	3.4	269	6	Q95KW7	Q95kw7 bos taurus
33	7	3.4	272	2	Q02304	Q02304 pseudomonas
34	7	3.4	276	5	P91894	P91894 arenicola m
35	7	3.4	277	10	Q9SKI1	Q9ski1 arabidopsis
36	7	3.4	291	15	Q9IUA3	Q9iua3 human immun
37	7	3.4	295	4	Q96BG3	Q96bg3 homo sapien
38	7	3.4	296	2	Q51756	Q51756 pseudomonas
39	7	3.4	299	2	Q9KZI2	Q9kzi2 streptomyce
40	7	3.4	299	12	Q9WRT4	Q9wrt4 macaca mula
41	7	3.4	301	16	Q9PLM7	Q9plm7 chlamydia m
42	7	3.4	303	11	Q9Z248	Q9z248 mus musculu
43	7	3.4	304	4	Q96HP9	Q96hp9 homo sapien
44	7	3.4	308	5	Q9NGT1	Q9ngt1 musca domes
45	7	3.4	318	10	Q9SJ37	Q9sj37 arabidopsis

ALIGNMENTS

RESULT 1
O95994
ID O95994 PRELIMINARY; PRT; 175 AA.
AC O95994;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SECRETED CEMENT GLAND PROTEIN XAG-2 HOMOLOG (ANTERIOR GRADIENT 2
DE (XENEPUS LAEVIS) HOMOLOG).
GN HAG-2/R OR HAG-2/C.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BREAST;
RX MEDLINE=99009231; PubMed=9790916;
RA Thompson D.A., Weigel R.J.;
RT "hAG-2, the human homologue of the Xenopus laevis cement gland gene
RT XAG-2, is coexpressed with estrogen receptor in breast cancer cell
RT lines.";
RL Biochem. Biophys. Res. Commun. 251:111-116(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Zhang J.S., Smith D.I.;
RT "Human homolog of XAG is differentially expressed in tumors.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF038451; AAC82614.1; -.
DR EMBL; AF007791; AAC77358.1; -.
DR EMBL; AF088867; AAF22484.1; -.
DR EMBL; BC015503; AAH15503.1; -.
SQ SEQUENCE 175 AA; 19979 MW; F271B1BD377BEE11 CRC64;

```
  Query Match             85.0%;  Score 175;  DB 4;  Length 175;
  Best Local Similarity   100.0%;  Pred. No. 2.4e-173;
  Matches  175;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      32 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 91
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 60

Qy      92 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 151
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 120

Qy     152 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 206
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 175


RESULT 2
O88312
ID   O88312      PRELIMINARY;      PRT;   175 AA.
AC   O88312;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   GOB-4 PROTEIN (ANTERIOR GRADIENT 2) (XENEPUS LAEVIS) (XENOPUS
DE   LAEVIS).
GN   AGR2 OR GOB-4 OR MAG-2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C; TISSUE=INTESTINE;
RA   Komiya T., Hirohashi S.;
RT   "Cloning of a gene, gob-4, which is expressed in intestinal goblet
RT   cells in mice.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SWISS WEBSTER;
RX   MEDLINE=99009231; PubMed=9790916;
RA   Thompson D.A., Weigel R.J.;
RT   "hAG-2, the human homologue of the Xenopus laevis cement gland gene
RT   XAG-2, is coexpressed with estrogen receptor in breast cancer cell
RT   lines.";
RL   Biochem. Biophys. Res. Commun. 251:111-116(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=PANCREAS;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [4]
     SEQUENCE FROM N.A.
RC   TISSUE=COLON, NORMAL. 5 MONTH OLD MALE MOUSE.;
RA   Strausberg R.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB016592; BAA32044.1; -.
DR   EMBL; AF044262; AAC72705.1; -.
DR   EMBL; AK007677; BAB25181.1; -.
DR   EMBL; BC013334; AAH13334.1; -.
DR   MGD; MGI:1344405; Agr2.
SQ   SEQUENCE   175 AA;  19920 MW;  ACC3CFE429B668CA CRC64;

  Query Match             16.5%;  Score 34;  DB 11;  Length 175;
  Best Local Similarity   100.0%;  Pred. No. 5.1e-27;
  Matches   34;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     126 KEIQKLAEQFVLLNLVYETTDKHLSPDGQYVPRI 159
           |||||||||||||||||||||||||||||||||
Db      95 KEIQKLAEQFVLLNLVYETTDKHLSPDGQYVPRI 128


RESULT 3
Q90Y05
ID   Q90Y05      PRELIMINARY;      PRT;   159 AA.
AC   Q90Y05;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CEMENT GLAND-SPECIFIC PROTEIN CGS.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Alexandrova E.M., Novoselov V.V., Zaraisky A.G.;
RT   "Three novel genes expressed in the anterior part of the Xenopus
RT   laevis embryo.";
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF314056; AAL26844.1; -.
SQ   SEQUENCE   159 AA;  18275 MW;  738D62284838B8EB CRC64;

  Query Match              4.9%;  Score 10;  DB 13;  Length 159;
  Best Local Similarity   100.0%;  Pred. No. 0.038;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      72 PQTLSRGWGD 81
           ||||||||||
Db      25 PQTLSRGWGD 34


RESULT 4
Q988W1
ID   Q988W1      PRELIMINARY;      PRT;   200 AA.
AC   Q988W1;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   MLR6569 PROTEIN.
GN   MLR6569.
OS   Rhizobium loti (Mesorhizobium loti).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Phyllobacteriaceae; Mesorhizobium.
OX   NCBI_TaxID=381;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MAFF303099;
RX   MEDLINE=21082930; PubMed=11214968;
RA   Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA   Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA   Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA   Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA   Takeuchi C., Yamada M., Tabata S.;
RT   "Complete genome structure of the nitrogen-fixing symbiotic bacterium
```

```
RT   Mesorhizobium loti.";
RL   DNA Res. 7:331-338(2000).
DR   EMBL; AP003009; BAB52836.1; -.
KW   Complete proteome.
SQ   SEQUENCE   200 AA;  21803 MW;  C9CA98E2F8CCF4DE CRC64;

  Query Match             3.9%;  Score 8;  DB 16;  Length 200;
  Best Local Similarity   100.0%;  Pred. No. 5.5;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     80 GDQLIWTQ 87
          ||||||||
Db     59 GDQLIWTQ 66


RESULT 5
Q62943
ID   Q62943      PRELIMINARY;      PRT;   315 AA.
AC   Q62943;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   TASTE BUD RECEPTOR PROTEIN TB 567.
GN   TB 567.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY;
RX   MEDLINE=97080538; PubMed=8921883;
RA   Thomas M.B., Haines S.L., Akeson R.A.;
RT   "Chemoreceptors expressed in taste, olfactory and male reproductive
RT   tissues.";
RL   Gene 178:1-5(1996).
DR   EMBL; U50948; AAC52910.1; -.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; UNKNOWN_1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   Receptor.
SQ   SEQUENCE   315 AA;  35556 MW;  650D03CD70558652 CRC64;

  Query Match             3.9%;  Score 8;  DB 11;  Length 315;
  Best Local Similarity   100.0%;  Pred. No. 8.1;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     41 LLLVALSY 48
          ||||||||
Db    142 LLLVALSY 149


RESULT 6
Q9RYQ8
ID   Q9RYQ8      PRELIMINARY;      PRT;   637 AA.
AC   Q9RYQ8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   ACYL-COA DEHYDROGENASE, PUTATIVE.
GN   DRA0250.
OS   Deinococcus radiodurans.
OC   Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX   NCBI_TaxID=1299;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R1;
RX   MEDLINE=20036896; PubMed=10567266;
RA   White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA   Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA   Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA   Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA   Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA   Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA   Fraser C.M.;
RT   "Genome sequence of the radioresistant bacterium Deinococcus
RT   radiodurans R1.";
RL   Science 286:1571-1577(1999).
DR   EMBL; AE001863; AAF12388.1; -.
DR   TIGR; DRA0250; -.
DR   InterPro; IPR001552; Acyl-CoA_dh.
DR   Pfam; PF00441; Acyl-CoA_dh; 2.
DR   Pfam; PF02770; Acyl-CoA_dh_M; 1.
KW   Complete proteome.
SQ   SEQUENCE   637 AA;  69706 MW;  EBCF21B96D1A4E18 CRC64;

  Query Match             3.9%;  Score 8;  DB 16;  Length 637;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    186 ADTALLLD 193
          ||||||||
Db    422 ADTALLLD 429


RESULT 7
Q960R8
ID   Q960R8      PRELIMINARY;      PRT;  1232 AA.
AC   Q960R8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   LD38671P.
GN   CG5634.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Y, CN BW SP;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA   Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA   Nunoo J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA   Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY051898; AAK93322.1; -.
SQ   SEQUENCE   1232 AA;  136912 MW;  CEF6003D6939A6DD CRC64;

  Query Match             3.9%;  Score 8;  DB 5;  Length 1232;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    179 RLYAYEPA 186
          ||||||||
Db    423 RLYAYEPA 430


RESULT 8
Q9VB20
ID   Q9VB20      PRELIMINARY;      PRT;  1284 AA.
AC   Q9VB20;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CG5634 PROTEIN.
GN   CG5634.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- SIMILARITY: CONTAINS 1 CUB DOMAIN.
DR   EMBL; AE003761; AAF56723.1; -.
DR   HSSP; P02468; 1KLO.
DR   FlyBase; FBgn0039528; CG5634.
DR   InterPro; IPR000859; CUB.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001798; Kelch.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR002165; Plexin_repeat.
DR   InterPro; IPR003659; PSI.
DR   Pfam; PF00431; CUB; 1.
DR   Pfam; PF01344; Kelch; 6.
DR   Pfam; PF00053; laminin_EGF; 2.
DR   Pfam; PF01437; Plexin_repeat; 1.
DR   SMART; SM00042; CUB; 1.
DR   SMART; SM00181; EGF; 1.
DR   SMART; SM00001; EGF_like; 1.
DR   SMART; SM00423; PSI; 4.
DR   PROSITE; PS01180; CUB; 2.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 1.
KW   EGF-like domain; Glycoprotein.
SQ   SEQUENCE   1284 AA;  142507 MW;  B3D559A1D22BF790 CRC64;

  Query Match             3.9%;  Score 8;  DB 5;  Length 1284;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     179 RLYAYEPA 186
           ||||||||
Db     475 RLYAYEPA 482
```

```
RESULT 9
Q91ER9
ID   Q91ER9      PRELIMINARY;      PRT;    73 AA.
AC   Q91ER9;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ORF136 SIMILAR TO XCGV ORF173.
GN   ORF136.
OS   Cydia pomonella granulosis virus (CpGV) (Cydia pomonella
OS   granulovirus).
OC   Viruses; dsDNA viruses, no RNA stage; Baculoviridae; Granulovirus.
OX   NCBI_TaxID=28289;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MEXICAN 1;
RX   MEDLINE=93188168; PubMed=8445726;
RA   Crook N.E., Clem R.J., Miller L.K.;
RT   "An apoptosis-inhibiting baculovirus gene with a zinc finger-like
RT   motif.";
RL   J. Virol. 67:2168-2174(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MEXICAN 1;
RX   MEDLINE=96207404; PubMed=8615018;
RA   Theilmann D.A., Chantler J.K., Stweart S., Flipsen H.T., Vlak J.M.,
RA   Crook N.E.;
RT   "Characterization of a highly conserved baculovirus structural protein
RT   that is specific for occlusion-derived virions.";
RL   Virology 218:148-158(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MEXICAN 1;
RX   MEDLINE=97380577; PubMed=9237352;
RA   Kang W., Crook N.E., Winstanley D., O'Reilly D.R.;
RT   "Complete sequence and transposon mutagenesis of the BamHI J fragment
RT   of Cydia pomonella granulosis virus.";
RL   Virus Genes 14:131-136(1997).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MEXICAN 1;
RX   MEDLINE=98418511; PubMed=9747739;
RA   Kang W., Tristem M., Maeda S., Crook N.E., O'Reilly D.R.;
RT   "Identification and characterization of the Cydia pomonella
RT   granulovirus cathepsin and chitinase genes.";
RL   J. Gen. Virol. 79:2283-2292(1998).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MEXICAN 1;
RA   Luque T., Finch R., Crook N., O'Reilly D.R., Winstanley D.;
RT   "The complete sequence of the Cydia pomonella granulovirus genome.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U53466; AAK70796.1; -.
SQ   SEQUENCE   73 AA;  8302 MW;  7CE181B2ED8B2F5B CRC64;

  Query Match             3.4%;  Score 7;  DB 12;  Length 73;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     131 LAEQFVL 137
           |||||||
Db      27 LAEQFVL 33


RESULT 10
O47957
ID   O47957      PRELIMINARY;      PRT;    80 AA.
AC   O47957;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   NADH DEHYDROGENASE SUBUNIT 6 (FRAGMENT).
```

```
OS   Phoxinus eos (northern redbelly dace).
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC   Cypriniformes; Cyprinidae; Phoxinus.
OX   NCBI_TaxID=42663;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96137174; PubMed=8564012;
RA   Toline C.A., Baker A.J.;
RT   "Mitochondrial DNA variation and population genetic structure of the
RT   northern redbelly dace (Phoxinus eos).";
RL   Mol. Ecol. 4:745-753(1995).
DR   EMBL; U34407; AAC01504.1; -.
DR   InterPro; IPR001457; Oxidored_q3.
DR   Pfam; PF00499; oxidored_q3; 1.
KW   Mitochondrion.
FT   NON_TER       1      1
FT   NON_TER      80     80
SQ   SEQUENCE   80 AA;  8042 MW;  D51E547824D59910 CRC64;

  Query Match             3.4%;  Score 7;  DB 8;  Length 80;
  Best Local Similarity   100.0%;  Pred. No. 28;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     40 FLLLVAL 46
          |||||||
Db      5 FLLLVAL 11


RESULT 11
Q56157
ID   Q56157      PRELIMINARY;      PRT;   101 AA.
AC   Q56157;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-JAN-1999 (TrEMBLrel. 09, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   C-5 ANTHRONE OXIDASE (FRAGMENT).
OS   Streptomyces venezuelae.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=54571;
RN   [1]
RP   SEQUENCE OF 1-93 FROM N.A.
RC   STRAIN=SP5230;
RX   MEDLINE=96146057; PubMed=8581159;
RA   Yang K., Han L., Ayer S.W., Vining L.C.;
RT   "Accumulation of the angucycline antibiotic rabelomycin after
RT   disruption of an oxygenase gene in the jadomycin B biosynthetic gene
RT   cluster of Streptomyces venezuelae.";
RL   Microbiology 142:0-0(0).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SP5230;
RA   Yang K.;
RL   Submitted (JUN-1995) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SP5230;
RA   Yang K.;
RL   Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SP5230;
RA   Han L.;
RL   Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U28382; AAC98518.1; -.
DR   InterPro; IPR001064; Crystallin.
DR   PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
FT   NON_TER     101    101
SQ   SEQUENCE   101 AA;  11457 MW;  21DC736567307512 CRC64;

  Query Match             3.4%;  Score 7;  DB 2;  Length 101;
  Best Local Similarity   100.0%;  Pred. No. 34;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    200 KLLKTEL 206
          |||||||
Db     95 KLLKTEL 101


RESULT 12
Q9Y8V3
ID   Q9Y8V3      PRELIMINARY;      PRT;   101 AA.
AC   Q9Y8V3;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   HYPOTHETICAL 10.3 KDA PROTEIN APE2531.
GN   APE2531.
OS   Aeropyrum pernix.
OC   Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
OC   Aeropyrum.
OX   NCBI_TaxID=56636;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K1;
RX   MEDLINE=99310339; PubMed=10382966;
RA   Kawarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA   Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA   Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA   Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA   Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA   Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT   "Complete genome sequence of an aerobic hyper-thermophilic
RT   crenarchaeon, Aeropyrum pernix K1.";
RL   DNA Res. 6:83-101(1999).
DR   EMBL; AP000064; BAA81547.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   101 AA;  10277 MW;  B8AF0E9B098E25B3 CRC64;

  Query Match             3.4%;  Score 7;  DB 17;  Length 101;
  Best Local Similarity   100.0%;  Pred. No. 34;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      5 AGTLSGS 11
          |||||||
Db     11 AGTLSGS 17


RESULT 13
Q9GZT3
ID   Q9GZT3      PRELIMINARY;      PRT;   109 AA.
AC   Q9GZT3;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   DC50 (DC23).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Xu X., Yang Y., Gao G., Xiao H., Chen Z., Han Z.;
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF271779; AAG44790.1; -.
DR   EMBL; AF253980; AAG44629.1; -.
DR   InterPro; IPR000504; RRM.
DR   Pfam; PF00076; rrm; 1.
DR   SMART; SM00360; RRM; 1.
DR   PROSITE; PS50102; RRM; 1.
SQ   SEQUENCE   109 AA;  12349 MW;  8BDCB6EA244F17B8 CRC64;

  Query Match             3.4%;  Score 7;  DB 4;  Length 109;
  Best Local Similarity   100.0%;  Pred. No. 36;
```

```
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      68 RPKLPQT 74
           |||||||
Db      95 RPKLPQT 101


RESULT 14
Q97SV7
ID   Q97SV7      PRELIMINARY;      PRT;   112 AA.
AC   Q97SV7;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CONSERVED DOMAIN PROTEIN.
GN   SP0207.
OS   Streptococcus pneumoniae.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1313;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TIGR4;
RX   MEDLINE=21357209; PubMed=11463916;
RA   Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA   Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA   Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA   Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA   Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA   McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA   Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA   Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT   "Complete genome sequence of a virulent isolate of Streptococcus
RT   pneumoniae.";
RL   Science 293:498-506(2001).
DR   EMBL; AE007335; AAK74387.1; -.
DR   TIGR; SP0207; -.
KW   Complete proteome.
SQ   SEQUENCE   112 AA;  13083 MW;  E21BB8790CDE54CE CRC64;

  Query Match             3.4%;  Score 7;  DB 16;  Length 112;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      50 LARDTTV 56
           |||||||
Db      60 LARDTTV 66


RESULT 15
Q9KXA4
ID   Q9KXA4      PRELIMINARY;      PRT;   114 AA.
AC   Q9KXA4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 12.4 KDA PROTEIN.
GN   H0152.
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7;
RX   MEDLINE=20198780; PubMed=10734605;
RA   Makino K., Yokoyama K., Kubota Y., Yutsudo C.H., Kimura S.,
RA   Kurokawa K., Ishii K., Hattori M., Tatsuno I., Abe H., Iida T.,
RA   Yamamoto K., Ohnishi M., Hayashi T., Yasunaga T., Honda T.,
RA   Sasakawa C., Shinagawa H.;
RT   "Complete nucleotide sequence of the prophage VT2-Sakai carrying the
RT   verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
RT   derived from the Sakai outbreak.";
RL   Genes Genet. Syst. 74:227-239(1999).
DR   EMBL; AP000422; BAA94180.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   114 AA;  12409 MW;  ABF77340011439AB CRC64;

  Query Match             3.4%;  Score 7;  DB 2;  Length 114;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     197 KALKLLK 203
           |||||||
Db     108 KALKLLK 114


Search completed: October  9, 2002, 16:44:07
Job time : 30 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 16:42:22 ; Search time 18 Seconds
(without alignments)
279.537 Million cell updates/sec

Title: US-09-674-266A-181
Perfect score: 206
Sequence: 1 RLSCAGTLSGSGPHPSRRLT.....DTALLLDNMKKALKLLKTEL 206

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 231628 seqs, 24425594 residues

Word size : 0

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	175	85.0	175	4	US-08-916-576B-2	Sequence 2, Appli
2	175	85.0	175	4	US-09-247-155-106	Sequence 106, App
3	59	28.6	131	4	US-09-247-155-174	Sequence 174, App
4	24	11.7	166	4	US-08-916-576B-6	Sequence 6, Appli
5	24	11.7	170	4	US-08-916-576B-8	Sequence 8, Appli
6	7	3.4	183	4	US-08-916-576B-7	Sequence 7, Appli
7	7	3.4	220	1	US-08-063-552-10	Sequence 10, Appl
8	7	3.4	220	5	PCT-US93-05704-10	Sequence 10, Appl
9	7	3.4	411	2	US-08-773-870-1	Sequence 1, Appli
10	7	3.4	442	3	US-08-993-359-28	Sequence 28, Appl
11	7	3.4	465	1	US-08-471-496-9	Sequence 9, Appli
12	7	3.4	465	2	US-08-894-840-9	Sequence 9, Appli
13	7	3.4	465	3	US-09-139-675-9	Sequence 9, Appli
14	7	3.4	470	1	US-08-471-496-2	Sequence 2, Appli
15	7	3.4	470	2	US-08-894-840-2	Sequence 2, Appli
16	7	3.4	470	3	US-09-139-675-2	Sequence 2, Appli
17	7	3.4	514	1	US-08-063-552-13	Sequence 13, Appl
18	7	3.4	514	5	PCT-US93-05704-13	Sequence 13, Appl
19	7	3.4	515	1	US-08-063-552-4	Sequence 4, Appli
20	7	3.4	515	5	PCT-US93-05704-4	Sequence 4, Appli
21	7	3.4	521	1	US-08-063-552-2	Sequence 2, Appli
22	7	3.4	521	5	PCT-US93-05704-2	Sequence 2, Appli
23	7	3.4	530	4	US-08-793-044-3	Sequence 3, Appli
24	6	2.9	21	1	US-08-944-133-13	Sequence 13, Appl
25	6	2.9	24	4	US-09-326-039-8	Sequence 8, Appli
26	6	2.9	29	2	US-08-194-981E-39	Sequence 39, Appl
27	6	2.9	29	2	US-08-194-981E-42	Sequence 42, Appl
28	6	2.9	29	2	US-08-194-981E-44	Sequence 44, Appl
29	6	2.9	44	1	US-08-530-010-31	Sequence 31, Appl
30	6	2.9	44	2	US-08-484-101B-31	Sequence 31, Appl
31	6	2.9	44	4	US-08-714-524D-31	Sequence 31, Appl
32	6	2.9	55	4	US-09-227-357-216	Sequence 216, App
33	6	2.9	63	4	US-09-326-039-2	Sequence 2, Appli
34	6	2.9	76	4	US-09-309-487-14	Sequence 14, Appl
35	6	2.9	76	4	US-09-309-487-16	Sequence 16, Appl
36	6	2.9	76	4	US-09-309-487-29	Sequence 29, Appl
37	6	2.9	92	4	US-09-309-487-21	Sequence 21, Appl
38	6	2.9	92	4	US-09-309-487-22	Sequence 22, Appl
39	6	2.9	110	4	US-08-858-207A-273	Sequence 273, App
40	6	2.9	116	3	US-08-545-809A-92	Sequence 92, Appl
41	6	2.9	116	3	US-08-545-809A-118	Sequence 118, App
42	6	2.9	116	3	US-08-545-809A-140	Sequence 140, App
43	6	2.9	117	3	US-08-545-809A-114	Sequence 114, App
44	6	2.9	118	3	US-08-545-809A-116	Sequence 116, App
45	6	2.9	118	3	US-08-545-809A-123	Sequence 123, App

ALIGNMENTS

RESULT 1
US-08-916-576B-2
; Sequence 2, Application US/08916576B
; Patent No. 6171816
; GENERAL INFORMATION:
; APPLICANT: YU, GUO-LIANG
; APPLICANT: DILLON, PATRICK J.
; APPLICANT: EBNER, REINHARD
; APPLICANT: ENDRESS, GREGORY A.
; TITLE OF INVENTION: NOVEL HUMAN GROWTH FACTORS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: US
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/916,576B
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,347
; FILING DATE: 23-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0500001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 175 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-916-576B-2

Query Match 85.0%; Score 175; DB 4; Length 175;
Best Local Similarity 100.0%; Pred. No. 1.3e-159;
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 32 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 91

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 60

Qy     92 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 151
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 120

Qy    152 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 206
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 175


RESULT 2
US-09-247-155-106
; Sequence 106, Application US/09247155A
; Patent No. 6312922
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, Jean-Baptiste
;  APPLICANT: Duclert, Aymeric
;  APPLICANT: Bougueleret, Lydie
;  TITLE OF INVENTION: Complementary DNAs
;  FILE REFERENCE: GENSET.021A
;  CURRENT APPLICATION NUMBER: US/09/247,155A
;  CURRENT FILING DATE: 1999-02-09
;  EARLIER APPLICATION NUMBER: 60/074,121
;  EARLIER FILING DATE: 1998-02-09
;  EARLIER APPLICATION NUMBER: 60/081,563
;  EARLIER FILING DATE: 1998-04-13
;  EARLIER APPLICATION NUMBER: 60/096,116
;  EARLIER FILING DATE: 1998-08-10
;  EARLIER APPLICATION NUMBER: 60/099,273
;  EARLIER FILING DATE: 1998-10-04
;  NUMBER OF SEQ ID NOS: 182
;  SOFTWARE: Patent.pm
; SEQ ID NO 106
;   LENGTH: 175
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SIGNAL
;   LOCATION: -20..-1
US-09-247-155-106

  Query Match             85.0%;  Score 175;  DB 4;  Length 175;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-159;
  Matches  175;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     32 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 91
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYEE 60

Qy     92 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 151
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAEQFVLLNLVYETTDKHLSP 120

Qy    152 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 206
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 DGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALLLDNMKKALKLLKTEL 175


RESULT 3
US-09-247-155-174
; Sequence 174, Application US/09247155A
; Patent No. 6312922
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, Jean-Baptiste
;  APPLICANT: Duclert, Aymeric
;  APPLICANT: Bougueleret, Lydie
;  TITLE OF INVENTION: Complementary DNAs
;  FILE REFERENCE: GENSET.021A
;  CURRENT APPLICATION NUMBER: US/09/247,155A
;  CURRENT FILING DATE: 1999-02-09
;  EARLIER APPLICATION NUMBER: 60/074,121
;  EARLIER FILING DATE: 1998-02-09
;  EARLIER APPLICATION NUMBER: 60/081,563
;  EARLIER FILING DATE: 1998-04-13
;  EARLIER APPLICATION NUMBER: 60/096,116
;  EARLIER FILING DATE: 1998-08-10
;  EARLIER APPLICATION NUMBER: 60/099,273
;  EARLIER FILING DATE: 1998-10-04
;  NUMBER OF SEQ ID NOS: 182
;  SOFTWARE: Patent.pm
; SEQ ID NO 174
;   LENGTH: 131
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SIGNAL
;   LOCATION: -20..-1
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: 40,41,43,60,70,76,82,86,105,107
;   OTHER INFORMATION: Xaa = any one of the twenty amino acids
US-09-247-155-174

  Query Match             28.6%;  Score 59;  DB 4;  Length 131;
  Best Local Similarity   100.0%;  Pred. No. 5.6e-49;
  Matches   59;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     32 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYE 90
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTLSRGWGDQLIWTQTYE 59


RESULT 4
US-08-916-576B-6
; Sequence 6, Application US/08916576B
; Patent No. 6171816
;  GENERAL INFORMATION:
;    APPLICANT:  YU, GUO-LIANG
;    APPLICANT:  DILLON, PATRICK J.
;    APPLICANT:  EBNER, REINHARD
;    APPLICANT:  ENDRESS, GREGORY A.
;    TITLE OF INVENTION:  NOVEL HUMAN GROWTH FACTORS
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  US
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/916,576B
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/024,347
;      FILING DATE:  23-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  STEFFE, ERIC K.
;      REGISTRATION NUMBER:  36,688
;      REFERENCE/DOCKET NUMBER:  1488.0500001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  166 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
```

```
;    MOLECULE TYPE:  protein
US-08-916-576B-6

  Query Match             11.7%;  Score 24;  DB 4;  Length 166;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-15;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    149 LSPDGQYVPRIMFVDPSLTVRADI 172
          ||||||||||||||||||||||||
Db    109 LSPDGQYVPRIMFVDPSLTVRADI 132


RESULT 5
US-08-916-576B-8
; Sequence 8, Application US/08916576B
; Patent No. 6171816
;  GENERAL INFORMATION:
;    APPLICANT:  YU, GUO-LIANG
;    APPLICANT:  DILLON, PATRICK J.
;    APPLICANT:  EBNER, REINHARD
;    APPLICANT:  ENDRESS, GREGORY A.
;    TITLE OF INVENTION:  NOVEL HUMAN GROWTH FACTORS
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  US
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/916,576B
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/024,347
;      FILING DATE:  23-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  STEFFE, ERIC K.
;      REGISTRATION NUMBER:  36,688
;      REFERENCE/DOCKET NUMBER:  1488.0500001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  170 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-916-576B-8

  Query Match             11.7%;  Score 24;  DB 4;  Length 170;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-15;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    149 LSPDGQYVPRIMFVDPSLTVRADI 172
          ||||||||||||||||||||||||
Db    112 LSPDGQYVPRIMFVDPSLTVRADI 135


RESULT 6
US-08-916-576B-7
; Sequence 7, Application US/08916576B
; Patent No. 6171816
;  GENERAL INFORMATION:
;    APPLICANT:  YU, GUO-LIANG
;    APPLICANT:  DILLON, PATRICK J.
;    APPLICANT:  EBNER, REINHARD
;    APPLICANT:  ENDRESS, GREGORY A.
;    TITLE OF INVENTION:  NOVEL HUMAN GROWTH FACTORS
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  US
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/916,576B
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/024,347
;      FILING DATE:  23-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  STEFFE, ERIC K.
;      REGISTRATION NUMBER:  36,688
;      REFERENCE/DOCKET NUMBER:  1488.0500001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  183 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-916-576B-7

  Query Match             3.4%;  Score 7;  DB 4;  Length 183;
  Best Local Similarity   100.0%;  Pred. No. 34;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    163 DPSLTVR 169
          |||||||
Db    140 DPSLTVR 146


RESULT 7
US-08-063-552-10
; Sequence 10, Application US/08063552
; Patent No. 5688936
;  GENERAL INFORMATION:
;    APPLICANT:  Edwards, Robert H
;    TITLE OF INVENTION:  Vesicle Membrane Transport Proteins
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sheldon & Mak
;      STREET:  225 South Lake Avenue, Ninth Floor
;      CITY:  Pasadena
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  91101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/063,552
;      FILING DATE:  19930514
;      CLASSIFICATION:  530
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Farber, Michael B
;      REGISTRATION NUMBER:  32,612
;      REFERENCE/DOCKET NUMBER:  9067-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (818) 796-4000
;      TELEFAX:  (818) 795-6321
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  220 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    HYPOTHETICAL:  NO
;    FRAGMENT TYPE:  internal
;    ORIGINAL SOURCE:
;      ORGANISM:  Rattus rattus
US-08-063-552-10

  Query Match             3.4%;  Score 7;  DB 1;  Length 220;
  Best Local Similarity   100.0%;  Pred. No. 40;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    189 ALLLDNM 195
          |||||||
Db     20 ALLLDNM 26


RESULT 8
PCT-US93-05704-10
; Sequence 10, Application PC/TUS9305704
;  GENERAL INFORMATION:
;    APPLICANT:  Edwards, Robert H
;    TITLE OF INVENTION:  Vesicle Membrane Transport Proteins
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sheldon & Mak
;      STREET:  225 South Lake Avenue, Ninth Floor
;      CITY:  Pasadena
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  91001
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/05704
;      FILING DATE:  19930611
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Farber, Michael B
;      REGISTRATION NUMBER:  32,612
;      REFERENCE/DOCKET NUMBER:  9067-1PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (818) 796-4000
;      TELEFAX:  (818) 795-6321
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  220 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    HYPOTHETICAL:  NO
;    FRAGMENT TYPE:  internal
;    ORIGINAL SOURCE:
;      ORGANISM:  Rattus rattus
PCT-US93-05704-10

  Query Match             3.4%;  Score 7;  DB 5;  Length 220;
  Best Local Similarity   100.0%;  Pred. No. 40;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    189 ALLLDNM 195
          |||||||
Db     20 ALLLDNM 26


RESULT 9
US-08-773-870-1
; Sequence 1, Application US/08773870
; Patent No. 5912143
;  GENERAL INFORMATION:
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Goli, Surya K.
;    TITLE OF INVENTION:  NOVEL HUMAN MAGE-LIKE PROTEIN
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/773,870
;      FILING DATE:  Herewith
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0179 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  411 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  Consensus
;      CLONE:  Consensus
US-08-773-870-1

  Query Match             3.4%;  Score 7;  DB 2;  Length 411;
  Best Local Similarity   100.0%;  Pred. No. 69;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    164 PSLTVRA 170
          |||||||
Db    132 PSLTVRA 138


RESULT 10
US-08-993-359-28
; Sequence 28, Application US/08993359A
; Patent No. 6039942
; GENERAL INFORMATION:
;  APPLICANT: Lassen, Soren F.
;  APPLICANT: Bech, Lisbeth
;  APPLICANT: Ohmann, Anders
;  APPLICANT: Breinholt, Jens
;  APPLICANT: Fuglsang, Claus C.
;  APPLICANT: Ostergaard, Peter R.
```

```
;  TITLE OF INVENTION: Phytase Polypeptides
;  FILE REFERENCE: 5383.500-US
;  CURRENT APPLICATION NUMBER: US/08/993,359A
;  CURRENT FILING DATE: 1997-12-18
;  EARLIER APPLICATION NUMBER: 1480/96
;  EARLIER FILING DATE: 1996-12-20
;  EARLIER APPLICATION NUMBER: 1481/96
;  EARLIER FILING DATE: 1996-12-20
;  EARLIER APPLICATION NUMBER: 0301/97
;  EARLIER FILING DATE: 1997-03-18
;  EARLIER APPLICATION NUMBER: 0529/97
;  EARLIER FILING DATE: 1997-05-07
;  EARLIER APPLICATION NUMBER: 1388/97
;  EARLIER FILING DATE: 1997-12-01
;  EARLIER APPLICATION NUMBER: 60/046,082
;  EARLIER FILING DATE: 1997-05-09
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 28
;    LENGTH: 442
;    TYPE: PRT
;    ORGANISM: Paxillus involtus
;    FEATURE:
;    NAME/KEY: SIGNAL
;    LOCATION: (1)...(19)
US-08-993-359-28

  Query Match             3.4%;  Score 7;  DB 3;  Length 442;
  Best Local Similarity   100.0%;  Pred. No. 73;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 RLSCAGT 7
          |||||||
Db    383 RLSCAGT 389


RESULT 11
US-08-471-496-9
; Sequence 9, Application US/08471496
; Patent No. 5798223
;  GENERAL INFORMATION:
;    APPLICANT:  LI, YI
;    APPLICANT:  CAO, LIANG
;    APPLICANT:  ROSEN, CRAIG
;    TITLE OF INVENTION:  HUMAN AMINE TRANSPORTER
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/471,496
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/US95/02645
;      FILING DATE:  01-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  STEFFE, ERIC K
;      REGISTRATION NUMBER:  36,688
;      REFERENCE/DOCKET NUMBER:  1488.0830001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  465 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-471-496-9

  Query Match             3.4%;  Score 7;  DB 1;  Length 465;
  Best Local Similarity   100.0%;  Pred. No. 77;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    189 ALLLDNM 195
          |||||||
Db     30 ALLLDNM 36


RESULT 12
US-08-894-840-9
; Sequence 9, Application US/08894840
; Patent No. 5859200
;  GENERAL INFORMATION:
;    APPLICANT:  LI, YI
;    APPLICANT:  CAO, LIANG
;    APPLICANT:  ROSEN, CRAIG A.
;    TITLE OF INVENTION:  HUMAN AMINE TRANSPORTER
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/894,840
;      FILING DATE:  29-AUG-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  STEFFE, ERIC K
;      REGISTRATION NUMBER:  36,688
;      REFERENCE/DOCKET NUMBER:  1488.0830000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  465 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-894-840-9

  Query Match             3.4%;  Score 7;  DB 2;  Length 465;
  Best Local Similarity   100.0%;  Pred. No. 77;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    189 ALLLDNM 195
          |||||||
Db     30 ALLLDNM 36


RESULT 13
US-09-139-675-9
; Sequence 9, Application US/09139675A
; Patent No. 6117426
; GENERAL INFORMATION:
;  APPLICANT: Li, Yi
```

```
;  APPLICANT: Cao, Liang
;  APPLICANT: Rosen, Craig
;  TITLE OF INVENTION: Human Amine Transporter
;  FILE REFERENCE: 1488.0830003
;  CURRENT APPLICATION NUMBER: US/09/139,675A
;  CURRENT FILING DATE: 1998-08-25
;  EARLIER APPLICATION NUMBER: WO PCT/US95/02645
;  EARLIER FILING DATE: 1995-03-01
;  EARLIER APPLICATION NUMBER: US 08/471,496
;  EARLIER FILING DATE: 1995-06-06
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
;   LENGTH: 465
;   TYPE: PRT
;   ORGANISM: Rattus sp.
US-09-139-675-9

  Query Match             3.4%;  Score 7;  DB 3;  Length 465;
  Best Local Similarity   100.0%;  Pred. No. 77;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     189 ALLLDNM 195
           |||||||
Db      30 ALLLDNM 36


RESULT 14
US-08-471-496-2
; Sequence 2, Application US/08471496
; Patent No. 5798223
;  GENERAL INFORMATION:
;    APPLICANT:  LI, YI
;    APPLICANT:  CAO, LIANG
;    APPLICANT:  ROSEN, CRAIG
;    TITLE OF INVENTION:  HUMAN AMINE TRANSPORTER
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/471,496
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/US95/02645
;      FILING DATE:  01-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  STEFFE, ERIC K
;      REGISTRATION NUMBER:  36,688
;      REFERENCE/DOCKET NUMBER:  1488.0830001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  470 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-471-496-2

  Query Match             3.4%;  Score 7;  DB 1;  Length 470;
  Best Local Similarity   100.0%;  Pred. No. 78;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     189 ALLLDNM 195
           |||||||
Db      30 ALLLDNM 36


RESULT 15
US-08-894-840-2
; Sequence 2, Application US/08894840
; Patent No. 5859200
;  GENERAL INFORMATION:
;    APPLICANT:  LI, YI
;    APPLICANT:  CAO, LIANG
;    APPLICANT:  ROSEN, CRAIG A.
;    TITLE OF INVENTION:  HUMAN AMINE TRANSPORTER
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/894,840
;      FILING DATE:  29-AUG-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  STEFFE, ERIC K
;      REGISTRATION NUMBER:  36,688
;      REFERENCE/DOCKET NUMBER:  1488.0830000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  470 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-894-840-2

  Query Match             3.4%;  Score 7;  DB 2;  Length 470;
  Best Local Similarity   100.0%;  Pred. No. 78;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     189 ALLLDNM 195
           |||||||
Db      30 ALLLDNM 36


Search completed: October  9, 2002, 16:45:11
Job time : 19 secs
```

```
OM of: US-09-674-266A-181 to: GenEmbl:*   out_format : pfs

Date: Oct 8, 2002 10:40 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlh
 -Q=/cgn2_1/USPTO_spool/US09674266/runat_08102002_093324_26161/app_query.fasta_1.271
 -DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US09674266_@CGN1_1_3973 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-674-266A-181
Query length: 206
Database: GenEmbl:*
Database sequences: 1797656
Database length: 1873333701
Search time (sec): 1921.540000


score_list:
Sequence            Strd Orig    ZScore    EScore Len  ! Documentation  ..
gb_pat:AX011612       + 1063.00 1896.06  3.0e-97   1020  ! AX011612 Sequence 8 from Paten
gb_pat:AX015056       + 1063.00 1896.06  3.0e-97   1020  ! AX015056 Sequence 265 from Pat
gb_pr:AF088867        + 1025.00 1828.39  1.8e-93   956  ! AF088867 Homo sapiens putative
gb_pr:AF038451        + 1002.00 1785.87  4.1e-91   1077  ! AF038451 Homo sapiens secreted
gb_pr:AF115926        + 1002.00 1781.36  7.4e-91   1701  ! AF115926 Homo sapiens XAG-2 ho
gb_pr:BC015503        +  995.00 1775.79  1.5e-90   836  ! BC015503 Homo sapiens, anterior
gb_pat:AR123828       +  991.00 1768.15  4.0e-90   875  ! AR123828 Sequence 1 from patent
gb_pat:AX067336       +  967.00 1726.17  8.7e-88   778  ! AX067336 Sequence 40 from Paten
gb_pat:AX341388       +  964.00 1723.45  1.2e-87   593  ! AX341388 Sequence 1635 from Pat
gb_pat:AR177367       +  961.00 1707.74  9.3e-87   1689  ! AR177367 Sequence 61 from pate
gb_pat:AX341084       +  957.00 1711.49  5.7e-87   557  ! AX341084 Sequence 1331 from Pat
gb_pat:AX341234       +  946.00 1690.93  8.0e-86   603  ! AX341234 Sequence 1481 from Pat
gb_pat:AR177410       +  914.00 1629.03  2.2e-82   940  ! AR177410 Sequence 149 from pate
gb_pat:AX231601       +  912.00 1630.85  1.8e-82   543  ! AX231601 Sequence 7 from Patent
gb_pr:AF007791        +  912.00 1626.57  3.1e-82   838  ! AF007791 Homo sapiens secreted
gb_pat:AX351468       -  884.00 1578.03  1.6e-79   699  ! AX351468 Sequence 215 from Pate
gb_pat:AX341953       +  876.00 1566.47  6.9e-79   525  ! AX341953 Sequence 2200 from Pat
gb_pat:AR144213       +  856.50 1527.36  1.0e-76   793  ! AR144213 Sequence 78 from paten
gb_pat:AR176400       +  856.50 1527.36  1.0e-76   793  ! AR176400 Sequence 78 from paten
gb_ro:AB016592        +  842.00 1501.74  2.8e-75   758  ! AB016592 Mus musculus mRNA for
gb_ro:AF044262        +  842.00 1501.44  2.9e-75   781  ! AF044262 Mus musculus secreted
gb_ro:BC013334        +  842.00 1501.18  3.0e-75   802  ! BC013334 Mus musculus, anterior
gb_pat:AX192616       +  702.00 1256.17  1.3e-61   409  ! AX192616 Sequence 183 from Pate
gb_pat:AX341186       -  691.00 1232.90  2.6e-60   583  ! AX341186 Sequence 1433 from Pat
gb_pat:AX341116       -  686.00 1223.83  8.3e-60   588  ! AX341116 Sequence 1363 from Pat
gb_pat:AR166839       +  683.00 1222.21  1.0e-59   401  ! AR166839 Sequence 32 from paten
gb_pat:AX192465       +  683.00 1222.21  1.0e-59   401  ! AX192465 Sequence 32 from Paten
gb_pat:AR123857       +  642.00 1148.89  1.2e-55   386  ! AR123857 Sequence 35 from paten
gb_pat:AX340984       -  627.00 1118.37  6.2e-54   554  ! AX340984 Sequence 1231 from Pat
gb_pat:AX341884       -  620.00 1105.97  3.1e-53   544  ! AX341884 Sequence 2131 from Pat
gb_pat:AX341524       -  616.00 1098.78  7.7e-53   544  ! AX341524 Sequence 1771 from Pat
gb_pat:AX340668       +  593.00 1058.03  1.4e-50   512  ! AX340668 Sequence 915 from Pate
gb_pat:AX341139       +  589.00 1051.09  3.5e-50   499  ! AX341139 Sequence 1386 from Pat
gb_pat:AX156194       -  580.00 1034.34  3.0e-49   529  ! AX156194 Sequence 51 from Paten
gb_pat:AX341229       -  576.00 1027.24  7.4e-49   524  ! AX341229 Sequence 1476 from Pat
gb_pat:AX231609       +  555.00  987.93  1.2e-46   614  ! AX231609 Sequence 7 from Patent
gb_pat:AX231718       +  555.00  987.93  1.2e-46   614  ! AX231718 Sequence 106 from Pate
gb_pat:AX193063       +  555.00  987.53  1.2e-46   640  ! AX193063 Sequence 630 from Pate
gb_pat:AR123830       +  555.00  986.20  1.4e-46   732  ! AR123830 Sequence 5 from patent
gb_ov:AF314056        +  553.00  982.18  2.4e-46   764  ! AF314056 Xenopus laevis cement
gb_pat:AX302898       -  547.00  979.95  3.2e-46   321  ! AX302898 Sequence 218 from P
gb_pat:AX322727       +  543.00  972.73  8.1e-46   322  ! AX322727 Sequence 2 from Pat
gb_pat:AX192553       +  516.00  924.86  3.8e-43   301  ! AX192553 Sequence 120 from P
gb_pat:AR123858       +  487.50  871.52  3.5e-40   373  ! AR123858 Sequence 36 from pa
gb_pat:AX341299       +  468.00  835.42  3.6e-38   415  ! AX341299 Sequence 1546 from

seq_name: gb_pat:AX011612

seq_documentation_block:
LOCUS       AX011612               1020 bp    DNA     linear   PAT 06-SEP-2000
DEFINITION  Sequence 8 from Patent WO9955858.
ACCESSION   AX011612
VERSION     AX011612.1  GI:9998136
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1020)
  AUTHORS   Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
            Pilarsky,C.
  TITLE     Human nucleic acid sequences obtained from pancreas tumor tissue
  JOURNAL   Patent: WO 9955858-A 8 04-NOV-1999;
            SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
            BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
            (DE); PILARSKY CHRISTIAN (DE)
FEATURES             Location/Qualifiers
     source          1. .1020
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      341 a    214 c    218 g    247 t
ORIGIN


alignment_scores:
           Quality: 1063.00              Length:    206
             Ratio:  5.160                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x AX011612   ..

 Align seg 1/1  to: AX011612  from: 1  to: 1020

       1 ArgLeuSerCysAlaGlyThrLeuSerGlySerGlyProHisProSerAr 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      57 AGACTCAGCTGTGCTGGCACACTCAGCGGCTCTGGACCGCATCCTAGCCG 106

      17 gArgLeuThrGlnGlyArgTrpValArgLysSerArgValAlaMetGluL 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     107 CCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAGAGTTGCCATGGAGA 156

      34 ysIleProValSerAlaPheLeuLeuLeuValAlaLeuSerTyrThrLeu 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     157 AAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCTCCTACACTCTG 206

      51 AlaArgAspThrThrValLysProGlyAlaLysLysAspThrLysAspSe 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     207 GCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGACACAAAGGACTC 256

      67 rArgProLysLeuProGlnThrLeuSerArgGlyTrpGlyAspGlnLeuI 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     257 TCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGGGTGACCAACTCA 306

      84 leTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerLysThrSerAsn 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     307 TCTGGACTCAGACATATGAAGAAGCTCTATATAAATCCAAGACAAGCAAC 356

     101 LysProLeuMetIleIleHisHisLeuAspGluCysProHisSerGlnAl 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     357 AAACCCTTGATGATTATTCATCACTTGGATGAGTGCCCACACAGTCAAGC 406

     117 aLeuLysLysValPheAlaGluAsnLysGluIleGlnLysLeuAlaGluG 134
```

```
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||
407 TTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCAGAAATTGGCAGAGC 456

134 lnPheValLeuLeuAsnLeuValTyrGluThrThrAspLysHisLeuSer 150
    ||||||||||||||||||||||||||||||||||||||||||||||||||
457 AGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTGACAAACACCTTTCT 506

151 ProAspGlyGlnTyrValProArgIleMetPheValAspProSerLeuTh 167
    ||||||||||||||||||||||||||||||||||||||||||||||||||
507 CCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTTGACCCATCTCTGAC 556

167 rValArgAlaAspIleThrGlyArgTyrSerAsnArgLeuTyrAlaTyrG 184
    ||||||||||||||||||||||||||||||||||||||||||||||||||
557 AGTTAGAGCCGATATCACTGGAAGATATTCAAACCGTCTCTATGCTTACG 606

184 luProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLysAlaLeuLys 200
    ||||||||||||||||||||||||||||||||||||||||||||||||||
607 AACCTGCAGATACAGCTCTGTTGCTTGACAACATGAAGAAAGCTCTCAAG 656

201 LeuLeuLysThrGluLeu 206
    ||||||||||||||||||
657 TTGCTGAAGACTGAATTG 674

seq_name: gb_pat:AX015056

seq_documentation_block:
LOCUS       AX015056                1020 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 265 from Patent WO9953040.
ACCESSION   AX015056
VERSION     AX015056.1  GI:10041195
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1020)
  AUTHORS   Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
            Pilarsky,C.
  TITLE     Human nucleic acid sequences from ovarian tumour tissue
  JOURNAL   Patent: WO 9953040-A 265 21-OCT-1999;
            SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
            BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
            (DE); PILARSKY CHRISTIAN (DE)
FEATURES             Location/Qualifiers
     source          1. .1020
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      341 a    214 c    218 g    247 t
ORIGIN


alignment_scores:
            Quality: 1063.00             Length:    206
              Ratio:  5.160               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x AX015056   ..

 Align seg 1/1  to: AX015056  from: 1  to: 1020

      1 ArgLeuSerCysAlaGlyThrLeuSerGlySerGlyProHisProSerAr 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     57 AGACTCAGCTGTGCTGGCACACTCAGCGGCTCTGGACCGCATCCTAGCCG 106

     17 gArgLeuThrGlnGlyArgTrpValArgLysSerArgValAlaMetGluL 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    107 CCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAGAGTTGCCATGGAGA 156

     34 ysIleProValSerAlaPheLeuLeuLeuValAlaLeuSerTyrThrLeu 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    157 AAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCTCCTACACTCTG 206

     51 AlaArgAspThrThrValLysProGlyAlaLysLysAspThrLysAspSe 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    207 GCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGACACAAAGGACTC 256

     67 rArgProLysLeuProGlnThrLeuSerArgGlyTrpGlyAspGlnLeuI 84
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    257 TCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGGGTGACCAACTCA 306

     84 leTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerLysThrSerAsn 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    307 TCTGGACTCAGACATATGAAGAAGCTCTATATAAATCCAAGACAAGCAAC 356

    101 LysProLeuMetIleIleHisHisLeuAspGluCysProHisSerGlnAl 117
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    357 AAACCCTTGATGATTATTCATCACTTGGATGAGTGCCCACACAGTCAAGC 406

    117 aLeuLysLysValPheAlaGluAsnLysGluIleGlnLysLeuAlaGluG 134
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    407 TTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCAGAAATTGGCAGAGC 456

    134 lnPheValLeuLeuAsnLeuValTyrGluThrThrAspLysHisLeuSer 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    457 AGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTGACAAACACCTTTCT 506

    151 ProAspGlyGlnTyrValProArgIleMetPheValAspProSerLeuTh 167
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    507 CCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTTGACCCATCTCTGAC 556

    167 rValArgAlaAspIleThrGlyArgTyrSerAsnArgLeuTyrAlaTyrG 184
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    557 AGTTAGAGCCGATATCACTGGAAGATATTCAAACCGTCTCTATGCTTACG 606

    184 luProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLysAlaLeuLys 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    607 AACCTGCAGATACAGCTCTGTTGCTTGACAACATGAAGAAAGCTCTCAAG 656

    201 LeuLeuLysThrGluLeu 206
        ||||||||||||||||||
    657 TTGCTGAAGACTGAATTG 674

seq_name: gb_pr:AF088867

seq_documentation_block:
LOCUS       AF088867                 956 bp    mRNA    linear   PRI 02-JAN-2000
DEFINITION  Homo sapiens putative secreted protein XAG mRNA, complete cds.
ACCESSION   AF088867
VERSION     AF088867.1  GI:6652811
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 956)
  AUTHORS   Zhang,J.S. and Smith,D.I.
  TITLE     Human homolog of XAG is differentially expressed in tumors
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 956)
  AUTHORS   Zhang,J.S. and Smith,D.I.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-AUG-1998) Pathology and Lab. Medicine, Mayo Clinic,
            200 SW 1st St., Rochester, MN 55905, USA
FEATURES             Location/Qualifiers
     source          1. .956
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             148. .675
                     /note="differentially expressed in tumors; similar to
                     Xenopus cement gland-specific secreted protein"
                     /codon_start=1
                     /product="putative secreted protein XAG"
                     /protein_id="AAF22484.1"
                     /db_xref="GI:6652812"
                     /translation="MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTL
```

```
                    SRGWGDQLIWTQTYEEALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAE
                    QFVLLNLVYETTDKHLSPDGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALL
                    LDNMKKALKLLKTEL"
BASE COUNT      310 a    202 c    202 g    242 t
ORIGIN


alignment_scores:
            Quality: 1025.00             Length:    206
              Ratio:  5.025               Gaps:      0
 Percent Similarity:  99.029   Percent Identity: 98.544

alignment_block:
 US-09-674-266A-181 x AF088867   ..

 Align seg 1/1  to: AF088867  from: 1  to: 956

       1 ArgLeuSerCysAlaGlyThrLeuSerGlySerGlyProHisProSerAr 17
         ||||||||||||||||||||||||    :::   ||||||||||||||||||
      56 AGACTCAGCTGTGCTGGCACACTCAGAAGCTT.GGACCGCATCCTAGCCG 104

      17 gArgLeuThrGlnGlyArgTrpValArgLysSerArgValAlaMetGluL 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     105 CCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAGAGTTGCCATGGAGA 154

      34 ysIleProValSerAlaPheLeuLeuLeuValAlaLeuSerTyrThrLeu 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     155 AAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCTCCTACACTCTG 204

      51 AlaArgAspThrThrValLysProGlyAlaLysLysAspThrLysAspSe 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     205 GCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGACACAAAGGACTC 254

      67 rArgProLysLeuProGlnThrLeuSerArgGlyTrpGlyAspGlnLeuI 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     255 TCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGGGTGACCAACTCA 304

      84 leTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerLysThrSerAsn 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     305 TCTGGACTCAGACATATGAAGAAGCTCTATATAAATCCAAGACAAGCAAC 354

     101 LysProLeuMetIleIleHisHisLeuAspGluCysProHisSerGlnAl 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     355 AAACCCTTGATGATTATTCATCACTTGGATGAGTGCCCACACAGTCAAGC 404

     117 aLeuLysLysValPheAlaGluAsnLysGluIleGlnLysLeuAlaGluG 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     405 TTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCAGAAATTGGCAGAGC 454

     134 lnPheValLeuLeuAsnLeuValTyrGluThrThrAspLysHisLeuSer 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     455 AGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTGACAAACACCTTTCT 504

     151 ProAspGlyGlnTyrValProArgIleMetPheValAspProSerLeuTh 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     505 CCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTTGACCCATCTCTGAC 554

     167 rValArgAlaAspIleThrGlyArgTyrSerAsnArgLeuTyrAlaTyrG 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     555 AGTTAGAGCCGATATCACTGGAAGATATTCAAATCGTCTCTATGCTTACG 604

     184 luProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLysAlaLeuLys 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     605 AACCTGCAGATACAGCTCTGTTGCTTGACAACATGAAGAAAGCTCTCAAG 654

     201 LeuLeuLysThrGluLeu 206
         |||||||||||||||||
     655 TTGCTGAAGACTGAATTG 672

seq_name: gb_pr:AF038451

seq_documentation_block:
LOCUS       AF038451               1077 bp    mRNA    linear   PRI 03-DEC-1998
DEFINITION  Homo sapiens secreted cement gland protein XAG-2 homolog (hAG-2/R)
            mRNA, complete cds.
ACCESSION   AF038451
VERSION     AF038451.1  GI:3779225
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1077)
  AUTHORS   Thompson,D.A. and Weigel,R.J.
  TITLE     hAG-2, the human homologue of the Xenopus laevis cement gland gene
            XAG-2, is coexpressed with estrogen receptor in breast cancer cell
            lines
  JOURNAL   Biochem. Biophys. Res. Commun. 251 (1), 111-116 (1998)
  MEDLINE   99009231
REFERENCE   2  (bases 1 to 1077)
  AUTHORS   Thompson,D.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1997) Surgery, Stanford University, 1201 Welch
            Road, MSLS Building, Room P228, Stanford, CA 94305, USA
FEATURES             Location/Qualifiers
     source          1. .1077
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="MCF7"
                     /tissue_type="breast"
                     /note="isolated from estrogen receptor-positive breast
                     adenocarcinoma cell line"
     gene            1. .1077
                     /gene="hAG-2/R"
     CDS             59. .586
                     /gene="hAG-2/R"
                     /note="similar to Homo sapiens protein hAG-2/C deposited
                     under GenBank Accession Number AF007791, and to Xenopus
                     laevis cement gland-specific proteins np77 and XAG-2
                     deposited under GenBank Accession Numbers U82110 and
                     AF025474, respectively"
                     /codon_start=1
                     /product="secreted cement gland protein XAG-2 homolog"
                     /protein_id="AAC82614.1"
                     /db_xref="GI:3779226"
                     /translation="MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTL
                     SRGWGDQLIWTQTYEEALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAE
                     QFVLLNLVYETTDKHLSPDGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALL
                     LDNMKKALKLLKTEL"
BASE COUNT      367 a    231 c    210 g    269 t
ORIGIN


alignment_scores:
            Quality: 1002.00             Length:    194
              Ratio:  5.165               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x AF038451   ..

 Align seg 1/1  to: AF038451  from: 1  to: 1077

      13 ProHisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerAr 29
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       2 CCGCATCCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAG 51

      29 gValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAlaL 46
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      52 AGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCC 101

      46 euSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLys 62
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     102 TCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAG 151

      63 AspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTr 79
```

```
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 152 GACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTG 201

  79 pGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysS 96
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 202 GGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAAT 251

  96 erLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCys 112
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 252 CCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGC 301

 113 ProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGl 129
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 302 CCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCA 351

 129 nLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrA 146
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 352 GAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTG 401

 146 spLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheVal 162
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 402 ACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTT 451

 163 AspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnAr 179
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 452 GACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAATCG 501

 179 gLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetL 196
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 502 TCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACATGA 551

 196 ysLysAlaLeuLysLeuLeuLysThrGluLeu 206
     |||||||||||||||||||||||||||||||
 552 AGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 583

seq_name: gb_pr:AF115926

seq_documentation_block:
LOCUS       AF115926               1701 bp    mRNA    linear   PRI 30-DEC-2001
DEFINITION  Homo sapiens XAG-2 homolog long protein (HPC8) mRNA, complete cds.
ACCESSION   AF115926
VERSION     AF115926.1  GI:17998664
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1701)
  AUTHORS   Zhang,J.S. and Smith,D.I.
  TITLE     Identification of human homolog of XAG-2 over-expressed in tumors
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1701)
  AUTHORS   Zhang,J.S. and Smith,D.I.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-DEC-1998) Laboratory of Medicine and Pathology, Mayo
            Clinic, 200 SW 1st Street, Rochester, MN 55905, USA
FEATURES             Location/Qualifiers
     source          1. .1701
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_type="prostate cancer"
     gene            1. .1701
                     /gene="HPC8"
     CDS             58. .585
                     /gene="HPC8"
                     /codon_start=1
                     /product="XAG-2 homolog long protein"
                     /protein_id="AAL54870.1"
                     /db_xref="GI:17998665"
                     /translation="MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTL
                     SRGWGDQLIWTQTYEEALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAE
                     QFVLLNLVYETTDKHLSPDGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALL
                     LDNMKKALKLLKTEL"

BASE COUNT      557 a    352 c    339 g    453 t
ORIGIN


alignment_scores:
            Quality: 1002.00             Length:    194
              Ratio:  5.165                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x AF115926   ..

 Align seg 1/1  to: AF115926  from: 1  to: 1701

      13 ProHisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerAr 29
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 CCGCATCCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAG 50

      29 gValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAlaL 46
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 AGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCC 100

      46 euSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLys 62
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 TCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAG 150

      63 AspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTr 79
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 GACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTG 200

      79 pGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysS 96
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 GGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAAT 250

      96 erLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCys 112
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 CCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGC 300

     113 ProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGl 129
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     301 CCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCA 350

     129 nLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrA 146
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     351 GAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTG 400

     146 spLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheVal 162
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     401 ACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTT 450

     163 AspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnAr 179
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     451 GACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAATCG 500

     179 gLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetL 196
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     501 TCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACATGA 550

     196 ysLysAlaLeuLysLeuLeuLysThrGluLeu 206
         |||||||||||||||||||||||||||||||
     551 AGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 582

seq_name: gb_pr:BC015503

seq_documentation_block:
LOCUS       BC015503               836 bp    mRNA    linear   PRI 29-OCT-2001
DEFINITION  Homo sapiens, anterior gradient 2 (Xenepus laevis) homolog, clone
            MGC:9112 IMAGE:3852448, mRNA, complete cds.
ACCESSION   BC015503
VERSION     BC015503.1  GI:15930124
KEYWORDS    MGC.
SOURCE      human.
  ORGANISM  Homo sapiens
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 836)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:       http://www-shgc.stanford.edu
            Contact:  (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 20 Row: o Column: 11
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 6652811.
FEATURES              Location/Qualifiers
     source           1. .836
                      /organism="Homo sapiens"
                      /db_xref="LocusID:10551"
                      /db_xref="taxon:9606"
                      /clone="MGC:9112 IMAGE:3852448"
                      /tissue_type="Colon, adenocarcinoma"
                      /clone_lib="NIH_MGC_65"
                      /lab_host="DH10B"
                      /note="Vector: pCMV-SPORT6"
     CDS              57. .584
                      /codon_start=1
                      /product="anterior gradient 2 (Xenepus laevis) homolog"
                      /protein_id="AAH15503.1"
                      /db_xref="GI:15930125"
                      /translation="MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTL
                      SRGWGDQLIWTQTYEEALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAE
                      QFVLLNLVYETTDKHLSPDGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALL
                      LDNMKKALKLLKTEL"
BASE COUNT      282 a    181 c    166 g    207 t
ORIGIN


alignment_scores:
            Quality:  995.00            Length:    193
              Ratio:  5.155               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x BC015503   ..

 Align seg 1/1  to: BC015503  from: 1  to: 836

      14 HisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerArgVa 30
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       3 CATCCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAGAGT 52

      30 lAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAlaLeuS 47
         |||||||||||||||||||||||||||||||||||||||||||||||||
      53 TGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCT 102

      47 erTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLysAsp 63
         |||||||||||||||||||||||||||||||||||||||||||||||||
     103 CCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGAC 152

      64 ThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTrpGl 80
         |||||||||||||||||||||||||||||||||||||||||||||||||
     153 ACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGGG 202

      80 yAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerL 97
         |||||||||||||||||||||||||||||||||||||||||||||||||
     203 TGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAATCCA 252

      97 ysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCysPro 113
         |||||||||||||||||||||||||||||||||||||||||||||||||
     253 AGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGCCCA 302

     114 HisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGlnLy 130
         |||||||||||||||||||||||||||||||||||||||||||||||||
     303 CACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCAGAA 352

     130 sLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrAspL 147
         |||||||||||||||||||||||||||||||||||||||||||||||||
     353 ATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTGACA 402

     147 ysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheValAsp 163
         |||||||||||||||||||||||||||||||||||||||||||||||||
     403 AACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTTGAC 452

     164 ProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnArgLe 180
         |||||||||||||||||||||||||||||||||||||||||||||||||
     453 CCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAACCGTCT 502

     180 uTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetLysL 197
         |||||||||||||||||||||||||||||||||||||||||||||||||
     503 CTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACATGAAGA 552

     197 ysAlaLeuLysLeuLeuLysThrGluLeu 206
         |||||||||||||||||||||||||||||
     553 AAGCTCTCAAGTTGCTGAAGACTGAATTG 581

seq_name: gb_pat:AR123828

seq_documentation_block:
LOCUS       AR123828                875 bp    DNA     linear   PAT 16-MAY-2001
DEFINITION  Sequence 1 from patent US 6171816.
ACCESSION   AR123828
VERSION     AR123828.1  GI:14109189
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 875)
  AUTHORS   Yu,G., Dillon,P.J., Ebner,R. and Endress,G.A.
  TITLE     Human XAG-1 polynucleotides and polypeptides
  JOURNAL   Patent: US 6171816-A 1 09-JAN-2001;
FEATURES              Location/Qualifiers
     source           1. .875
                      /organism="unknown"
BASE COUNT      283 a    189 c    180 g    223 t
ORIGIN


alignment_scores:
            Quality:  991.00            Length:    194
              Ratio:  5.135               Gaps:      0
 Percent Similarity:  99.485   Percent Identity: 99.485

alignment_block:
 US-09-674-266A-181 x AR123828   ..

 Align seg 1/1  to: AR123828  from: 1  to: 875

      13 ProHisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerAr 29
         |||  ||||||||||||||||||||||||||||||||||||||||||||||
      14 CCGATTCCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAG 63

      29 gValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAlaL 46
         |||||||||||||||||||||||||||||||||||||||||||||||||
```

```
     64 AGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCC 113

     46 euSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLys 62
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    114 TCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAG 163

     63 AspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTr 79
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    164 GACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTG 213

     79 pGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysS 96
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    214 GGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAAT 263

     96 erLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCys 112
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    264 CCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGC 313

    113 ProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGl 129
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    314 CCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCA 363

    129 nLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrA 146
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    364 GAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTG 413

    146 spLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheVal 162
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    414 ACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTT 463

    163 AspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnAr 179
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    464 GACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAATCG 513

    179 gLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetL 196
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    514 TCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACATGA 563

    196 ysLysAlaLeuLysLeuLeuLysThrGluLeu 206
        |||||||||||||||||||||||||||||||
    564 AGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 595

seq_name: gb_pat:AX067336

seq_documentation_block:
LOCUS       AX067336                778 bp    DNA     linear   PAT 24-JAN-2001
DEFINITION  Sequence 40 from Patent WO0078960.
ACCESSION   AX067336
VERSION     AX067336.1  GI:12544960
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 778)
  AUTHORS   Yuqiu,J. and Mitcham,J.L.
  TITLE     Compositions and methods for the therapy and diagnosis of breast
            cancer
  JOURNAL   Patent: WO 0078960-A 40 28-DEC-2000;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .778
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     misc_feature    1. .778
                     /note="n = A,T,C or G"
BASE COUNT      245 a    175 c    158 g    193 t      7 others
ORIGIN


alignment_scores:
            Quality:  967.00             Length:    194
              Ratio:  5.036                Gaps:      0
 Percent Similarity:  98.969   Percent Identity: 98.969

alignment_block:
 US-09-674-266A-181 x AX067336   ..

 Align seg 1/1  to: AX067336  from: 1  to: 778

     13 ProHisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerAr 29
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      2 CCGCATCCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAG 51

     29 gValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAlaL 46
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     52 AGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCC 101

     46 euSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLys 62
        ||||||||||||||||||||   ||||||||||||||||||||||||||
    102 TCTCCTACACTCTGGCCAGASATACCACAGTCAAACCTGGAGCCAAAAAG 151

     63 AspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTr 79
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    152 GACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTG 201

     79 pGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysS 96
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    202 GGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAAT 251

     96 erLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCys 112
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    252 CCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGC 301

    113 ProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGl 129
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    302 CCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCA 351

    129 nLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrA 146
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    352 GAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTG 401

    146 spLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheVal 162
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    402 ACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTT 451

    163 AspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnAr 179
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    452 GACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAATCG 501

    179 gLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetL 196
        ||||||||||||||||||||||   ||||||||||||||||||||||||
    502 TCTCTATGCTTACGAACCTGCAGATCA.GCTCTGTTGCTTGACAACATGA 550

    196 ysLysAlaLeuLysLeuLeuLysThrGluLeu 206
        |||||||||||||||||||||||||||||||
    551 AGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 582

seq_name: gb_pat:AX341388

seq_documentation_block:
LOCUS       AX341388                593 bp    DNA     linear   PAT 10-JAN-2002
DEFINITION  Sequence 1635 from Patent WO0196388.
ACCESSION   AX341388
VERSION     AX341388.1  GI:18137370
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Jiang,Y., Harlocker,S.L. and Secrist,H.
  TITLE     Compositions and methods for the therapy and diagnosis of colon
            cancer
  JOURNAL   Patent: WO 0196388-A 1635 20-DEC-2001;
            CORIXA CORPORATION (US)
```

```
FEATURES             Location/Qualifiers
     source          1. .593
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      188 a    145 c    120 g    138 t      2 others
ORIGIN


alignment_scores:
            Quality:  964.00             Length:    189
              Ratio:  5.128                Gaps:      0
 Percent Similarity:  99.471   Percent Identity: 99.471

alignment_block:
 US-09-674-266A-181 x AX341388   ..

 Align seg 1/1  to: AX341388  from: 1  to: 593

      18 ArgLeuThrGlnGlyArgTrpValArgLysSerArgValAlaMetGluLy 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       3 CGACTCACACAAGGCAGGTGGGTGAGGAAATCCAGAGTTGCCATGGAGAA 52

      34 sIleProValSerAlaPheLeuLeuLeuValAlaLeuSerTyrThrLeuA 51
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      53 AATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCTCCTACACTCTGG 102

      51 laArgAspThrThrValLysProGlyAlaLysLysAspThrLysAspSer 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     103 CCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGACACAAAGGACTCT 152

      68 ArgProLysLeuProGlnThrLeuSerArgGlyTrpGlyAspGlnLeuIl 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     153 CGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGGGTGACCAACTCAT 202

      84 eTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerLysThrSerAsnL 101
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     203 CTGGACTCAGACATATGAAGAAGCTCTATATAAATCCAAGACAAGCAACA 252

     101 ysProLeuMetIleIleHisHisLeuAspGluCysProHisSerGlnAla 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     253 AACCCTTGATGATTATTCATCACTTGGATGAGTGCCCACACAGTCAAGCT 302

     118 LeuLysLysValPheAlaGluAsnLysGluIleGlnLysLeuAlaGluGl 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     303 TTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCAGAAATTGGCAGAGCA 352

     134 nPheValLeuLeuAsnLeuValTyrGluThrThrAspLysHisLeuSerP 151
         |||||||||| ||||||||||||||||||||||||||||||||||||||||
     353 GTTTGTCCTCCCCAATCTGGTTTATGAAACAACTGACAAACACCTTTCTC 402

     151 roAspGlyGlnTyrValProArgIleMetPheValAspProSerLeuThr 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     403 CTGATGGCCAGTATGTCCCCAGGATTATGTTTGTTGACCCATCTCTGACA 452

     168 ValArgAlaAspIleThrGlyArgTyrSerAsnArgLeuTyrAlaTyrGl 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     453 GTTAGAGCCGATATCACTGGAAGATATTCAAATCGTCTCTATGCTTACGA 502

     184 uProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLysAlaLeuLysL 201
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     503 ACCTGCAGATACAGCTCTGTTGCTTGACAACATGAAGAAAGCTCTCAAGT 552

     201 euLeuLysThrGluLeu 206
         |||||||||||||||||
     553 TGCTGAAGACTGAATTG 569

seq_name: gb_pat:AR177367

seq_documentation_block:
LOCUS       AR177367               1689 bp    DNA     linear   PAT 17-DEC-2001
DEFINITION  Sequence 61 from patent US 6312922.
ACCESSION   AR177367
VERSION     AR177367.1  GI:17919722
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1689)
  AUTHORS   Edwards,J.-B.Dumas.Milne., Duclert,A. and Bougueleret,L.
  TITLE     Complementary DNAs
  JOURNAL   Patent: US 6312922-A 61 06-NOV-2001;
FEATURES             Location/Qualifiers
     source          1. .1689
                     /organism="unknown"
BASE COUNT      552 a    350 c    335 g    452 t
ORIGIN


alignment_scores:
            Quality:  961.00             Length:    195
              Ratio:  5.085                Gaps:      1
 Percent Similarity:  96.923   Percent Identity: 96.923

alignment_block:
 US-09-674-266A-181 x AR177367   ..

 Align seg 1/1  to: AR177367  from: 1  to: 1689

      12 GlyProHisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSe 28
         ||||||||||||||||||||||||||||||||
       9 GGACCGCATCCTAGCCGCCGACTCACACAAGGC..................  41

      28 rArgValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValA 45
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      42 .AGAGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGG 90

      45 laLeuSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLys 61
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      91 CCCTCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAA 140

      62 LysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGl 78
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     141 AAGGACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGG 190

      78 yTrpGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrL 95
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     191 TTGGGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATA 240

      95 ysSerLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGlu 111
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     241 AATCCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAG 290

     112 CysProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIl 128
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     291 TGCCCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAAT 340

     128 eGlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrT 145
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     341 CCAGAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAA 390

     145 hrAspLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPhe 161
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     391 CTGACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTT 440

     162 ValAspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAs 178
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     441 GTTGACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAA 490

     178 nArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnM 195
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     491 TCGTCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACA 540

     195 etLysLysAlaLeuLysLeuLeuLysThrGluLeu 206
         |||||||||||||||||||||||||||||||||||
     541 TGAAGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 575
```

```
seq_name: gb_pat:AX341084

seq_documentation_block:
LOCUS       AX341084                557 bp    DNA     linear   PAT 10-JAN-2002
DEFINITION  Sequence 1331 from Patent WO0196388.
ACCESSION   AX341084
VERSION     AX341084.1  GI:18137066
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Jiang,Y., Harlocker,S.L. and Secrist,H.
  TITLE     Compositions and methods for the therapy and diagnosis of colon
            cancer
  JOURNAL   Patent: WO 0196388-A 1331 20-DEC-2001;
      .     CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .557
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      169 a    143 c    113 g    130 t      2 others
ORIGIN


alignment_scores:
            Quality:  957.00             Length:    184
              Ratio:  5.201                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x AX341084   ..

 Align seg 1/1  to: AX341084  from: 1  to: 557

      13 ProHisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerAr 29
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
       1 CCGCATCCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAG 50

      29 gValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAlaL 46
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      51 AGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCC 100

      46 euSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLys 62
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     101 TCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAG 150

      63 AspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTr 79
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     151 GACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTG 200

      79 pGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysS 96
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     201 GGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAAT 250

      96 erLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCys 112
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     251 CCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGC 300

     113 ProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGl 129
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     301 CCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCA 350

     129 nLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrA 146
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     351 GAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTG 400

     146 spLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheVal 162
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     401 ACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTT 450

     163 AspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnAr 179
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     451 GACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAATCG 500

     179 gLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetL 196
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     501 TCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACATGA 550

     196 ys 196
         ||
     551 AG 552

seq_name: gb_pat:AX341234

seq_documentation_block:
LOCUS       AX341234                603 bp    DNA     linear   PAT 10-JAN-2002
DEFINITION  Sequence 1481 from Patent WO0196388.
ACCESSION   AX341234
VERSION     AX341234.1  GI:18137216
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Jiang,Y., Harlocker,S.L. and Secrist,H.
  TITLE     Compositions and methods for the therapy and diagnosis of colon
            cancer
  JOURNAL   Patent: WO 0196388-A 1481 20-DEC-2001;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .603
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      187 a    153 c    116 g    144 t      3 others
ORIGIN


alignment_scores:
            Quality:  946.00             Length:    193
              Ratio:  5.059                Gaps:      1
 Percent Similarity:  96.891   Percent Identity: 96.373

alignment_block:
 US-09-674-266A-181 x AX341234   ..

 Align seg 1/1  to: AX341234  from: 1  to: 603

      14 HisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerArgVa 30
         ||||||||||||||||||||||||||||                  |||||
       3 CATCCTAGCCGCCGACTCACACAAGGC..................AGAGT 34

      30 lAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAlaLeuS 47
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      35 TGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCT 84

      47 erTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLysAsp 63
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      85 CCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGAC 134

      64 ThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTrpGl 80
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     135 ACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGGG 184

      80 yAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerL 97
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     185 TGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAATCCA 234

      97 ysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCysPro 113
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     235 AGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGCCCA 284

     114 HisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGlnLy 130
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
     285 CACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCAGAA 334

     130 sLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrAspL 147
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     335 ATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTGACA 384

     147 ysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheValAsp 163
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     385 AACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTTGAC 434

     164 ProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnArgLe 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     435 CCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAATCGTCT 484

     180 uTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetLysL 197
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     485 CTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACATGAAGA 534

     197 ysAlaLeuLysLeuLeuLysThrGluLeu 206
         |||||||||||||||||||||||:::
     535 AAGCTCTCAAGTTGCTGAAGACTGAAATT 563

seq_name: gb_pat:AR177410

seq_documentation_block:
LOCUS       AR177410                940 bp    DNA     linear   PAT 17-DEC-2001
DEFINITION  Sequence 149 from patent US 6312922.
ACCESSION   AR177410
VERSION     AR177410.1  GI:17919765
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 940)
  AUTHORS   Edwards,J.-B.Dumas.Milne., Duclert,A. and Bougueleret,L.
  TITLE     Complementary DNAs
  JOURNAL   Patent: US 6312922-A 149 06-NOV-2001;
FEATURES             Location/Qualifiers
     source          1. .940
                     /organism="unknown"
BASE COUNT      304 a    202 c    187 g    233 t     14 others
ORIGIN


alignment_scores:
            Quality:  914.00             Length:    207
              Ratio:  4.811                Gaps:      2
 Percent Similarity:  91.787   Percent Identity: 90.821

alignment_block:
 US-09-674-266A-181 x AR177410   ..

 Align seg 1/1  to: AR177410  from: 1  to: 940

       1 ArgLeuSerCysAlaGlyThrLeuSerGlySerGlyProHisProSerAr 17
         |||||||||||||||||||||   :::   ||||||||||||||||||
     103 AGACTCAGCTGTGCTGGCACACTCAGAAGCTT.GGACCGCATCCTAGCCG 151

      17 gArgLeuThrGlnGlyArgTrpValArgLysSerArgValAlaMetGluL 34
         ||||||||||||||||                ||||||||||||||||
     152 CCGACTCACACAAGGC................AGAGTTGCCATGGAAA 183

      34 ysIleProValSerAlaPheLeuLeuLeuValAlaLeuSerTyrThrLeu 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     184 AAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCTCCTACACTCTG 233

      51 AlaArgAspThrThrValLysProGlyAlaLysLysAspThrLysAspSe 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     234 GCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGACACAAAGGACTC 283

      67 rArgProLysLeuProGlnThrLeuSerArgGlyTrpGlyAspGlnLeuI 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     284 TCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGGGTGACCAACTCA 333

      84 leTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerLysThrSerAsn 100
         ||||||||||||||||||||   :::|||   ||||||||||||||||||
     334 TCTGGACACARACATATGAARAARCTCTATWTAAATCCAARACAAGCAAC 383

     101 LysProLeuMetIleIleHisHisLeuAspGluCysProHisSerGlnAl 117
         |||||||||||||||||||||||||||||   ||||||||||||||||||
     384 AAACCCTTGATGATTATTCATCACTTGGATGADTGCCCACACAGTCAAGC 433

     117 aLeuLysLysValPheAlaGluAsnLysGluIleGlnLysLeuAlaGluG 134
         ||||||||||   ||||||||||||||   ||||||||||||||   |
     434 TTTAAAAAAAAKTGTTTGCTGAAAATAAARAAATCCAGAAATTGGCARANC 483

     134 lnPheValLeuLeuAsnLeuValTyrGluThrThrAspLysHisLeuSer 150
         ||||||||   |||||||||||||||||||||||||||||||||||||||
     484 AGTTTGTCCYCCTCAATCTGGTTTATGAAACAACTGACAAACACCTTTCT 533

     151 ProAspGlyGlnTyrVal.ProArgIleMetPheValAspProSerLeuT 167
         ||||||||||||||||   ||||||||||||||||||||||||||||||
     534 CCTGATGGCCAATATKTCCCCCMGGATTATGTTTGTTGACCCATCTCTGA 583

     167 hrValArgAlaAspIleThrGlyArgTyrSerAsnArgLeuTyrAlaTyr 183
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     584 CAGTTAGAGCCGATATCACTGGAAGATATTCAAAYCGTCTCTATGCTTAC 633

     184 GluProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLysAlaLeuLy 200
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     634 GAACCTGCAGATACAGCTCTGTTGCTTGACAACATGAAGAAAGCTCTCAA 683

     200 sLeuLeuLysThrGluLeu 206
         ||||||||||||||||||
     684 GTTGCTGAAGACTGAATTG 702

seq_name: gb_pat:AX231601

seq_documentation_block:
LOCUS       AX231601                543 bp    DNA     linear   PAT 11-SEP-2001
DEFINITION  Sequence 7 from Patent WO0163290.
ACCESSION   AX231601
VERSION     AX231601.1  GI:15592500
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 543)
  AUTHORS   Boyd,R.S., Stamps,A.C., Terrett,J.A. and Tyson,K.L.
  TITLE     Bcmp-7 as marker for diagnosis of breast cancer
  JOURNAL   Patent: WO 0163290-A 7 30-AUG-2001;
            Oxford GlycoSciences (UK) Limited (GB)
FEATURES             Location/Qualifiers
     source          1. .543
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             11. .538
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAC69703.1"
                     /db_xref="GI:15592501"
                     /translation="MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTL
                     SRGWGDQLIWTQTYEEALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAE
                     QFVLLNLVYETTDKHLSPDGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALL
                     LDNMKKALKLLKTEL"
BASE COUNT      170 a    135 c    107 g    131 t
ORIGIN


alignment_scores:
            Quality:  912.00             Length:    178
              Ratio:  5.124                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
```

```
US-09-674-266A-181 x AX231601   ..

Align seg 1/1  to: AX231601  from: 1  to: 543

     29 ArgValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAl 45
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      2 AGAGTTGCCATGGAGAAAATCCCAGTGTCAGCATTCTTGCTCCTTGTGGC 51

     45 aLeuSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysL 62
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     52 CCTCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAA 101

     62 ysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGly 78
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    102 AGGACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGT 151

     79 TrpGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLy 95
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    152 TGGGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAA 201

     95 sSerLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluC 112
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    202 ATCCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGT 251

    112 ysProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIle 128
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    252 GCCCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATC 301

    129 GlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrTh 145
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    302 CAGAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAAC 351

    145 rAspLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheV 162
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    352 TGACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTG 401

    162 alAspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsn 178
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    402 TTGACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAAC 451

    179 ArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMe 195
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    452 CGTCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACAT 501

    195 tLysLysAlaLeuLysLeuLeuLysThrGluLeu 206
        |||||||||||||||||||||||||||||||||
    502 GAAGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 535

seq_name: gb_pr:AF007791

seq_documentation_block:
LOCUS       AF007791                838 bp    mRNA    linear   PRI 14-NOV-1998
DEFINITION  Homo sapiens secreted cement gland protein XAG-2 homolog (hAG-2/C)
            mRNA, complete cds.
ACCESSION   AF007791
VERSION     AF007791.1  GI:3779196
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 838)
  AUTHORS   Thompson,D.A. and Weigel,R.J.
  TITLE     hAG-2, the human homologue of the Xenopus laevis cement gland gene
            XAG-2, is coexpressed with estrogen receptor in breast cancer cell
            lines
  JOURNAL   Biochem. Biophys. Res. Commun. 251 (1), 111-116 (1998)
  MEDLINE   99009231
REFERENCE   2  (bases 1 to 838)
  AUTHORS   Thompson,D.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUN-1997) Surgery, Stanford, 1201 Welch Road, MSLS
            Building, Room P228, Stanford, CA 94305, USA

FEATURES             Location/Qualifiers
     source          1. .838
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="MCF7; estrogen receptor-expressing"
                     /cell_type="adenocarcinoma"
                     /tissue_type="breast"
     gene            1. .838
                     /gene="hAG-2/C"
     CDS             11. .538
                     /gene="hAG-2/C"
                     /note="similar to Homo sapiens protein hAG-2/R deposited
                     in GenBank Accession Number AF038451, and to Xenopus
                     laevis cement gland proteins np77 deposited in GenBank
                     Accession Number U82110 and XAG-2 deposited in GenBank
                     Accession Number AF025474"
                     /codon_start=1
                     /product="secreted cement gland protein XAG-2 homolog"
                     /protein_id="AAC77358.1"
                     /db_xref="GI:3779197"
                     /translation="MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTL
                     SRGWGDQLIWTQTYEEALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAE
                     QFVLLNLVYETTDKHLSPDGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALL
                     LDNMKKALKLLKTEL"
BASE COUNT      291 a    170 c    160 g    217 t
ORIGIN


alignment_scores:
           Quality:  912.00            Length:    178
             Ratio:  5.124               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x AF007791   ..

 Align seg 1/1  to: AF007791  from: 1  to: 838

      29 ArgValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAl 45
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       2 AGAGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGC 51

      45 aLeuSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysL 62
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      52 CCTCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAA 101

      62 ysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGly 78
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     102 AGGACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGT 151

      79 TrpGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLy 95
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     152 TGGGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAA 201

      95 sSerLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluC 112
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     202 ATCCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGT 251

     112 ysProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIle 128
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     252 GCCCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATC 301

     129 GlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrTh 145
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     302 CAGAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAAC 351

     145 rAspLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheV 162
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     352 TGACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTG 401

     162 alAspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsn 178
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     402 TTGACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAAT 451
```

```
179 ArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMe 195
    ||||||||||||||||||||||||||||||||||||||||||||||||||
452 CGTCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACAT 501

195 tLysLysAlaLeuLysLeuLeuLysThrGluLeu 206
    |||||||||||||||||||||||||||||||||
502 GAAGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 535
```

```
OM of: US-09-674-266A-181 to: N_Geneseq_032802:*   out_format : pfs

Date: Oct 8, 2002 10:07 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlh
 -Q=/cgn2_1/USPTO_spool/US09674266/runat_08102002_093325_26197/app_query.fasta_1.271
 -DB=N_Geneseq_032802 -QFMT=fastap -SUFFIX=p2n.rng -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US09674266_@CGN1_1_0 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-674-266A-181
Query length: 206
Database: N_Geneseq_032802:*
Database sequences: 1736436
Database length: 858457221
Search time (sec): 213.400000


score_list:
Sequence          Strd Orig    ZScore    EScore Len  ! Documentation  ..
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAZ52865 + 1063.00 2063.38  1.4e-106  10
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAC58380 + 1025.00 1989.96  1.7e-102  92
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAZ77486 + 1018.50 1976.12  1.0e-101  10
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:AAV29047 + 1008.00 1957.27  1.1e-100  86
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH35026 + 1002.00 1944.58  5.8e-100  9
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:AAV19155 +  991.00 1923.81  8.3e-99   87
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAF63314 +  991.00 1923.81  8.3e-99   8
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:AAV59320 +  987.00 1915.89  2.3e-98   88
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAF44884 +  967.00 1878.00  2.9e-96   7
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAZ40803 +  961.00 1857.91  3.9e-95   16
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAZ40846 +  914.00 1772.02  2.4e-90   94
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAS13480 +  912.00 1774.00  1.8e-90   5
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:AAV29048 +  899.00 1748.86  4.6e-89   52
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAZ24578 +  856.50 1661.08  3.6e-84   79
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAC65817 +  856.50 1661.08  3.6e-84   79
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAC00115 +  768.00 1493.17  8.0e-75   46
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAA77903 +  702.00 1365.17  1.1e-67   40
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAI28641 +  702.00 1365.17  1.1e-67   4
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:AAV59329 +  688.00 1335.03  5.2e-66   52
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAA77753 +  683.00 1328.12  1.3e-65   40
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAI28491 +  683.00 1328.12  1.3e-65   4
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAZ42286 +  672.00 1303.75  2.8e-64   52
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAL23173 -  665.50 1284.69  3.3e-63   9
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAX41145 +  660.00 1283.09  4.0e-63   39
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:AAV19188 +  642.00 1248.11  3.6e-61   38
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAF63347 +  642.00 1248.11  3.6e-61   3
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAC65978 +  631.00 1226.05  6.1e-60   40
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAL19619 +  589.50 1145.72  1.8e-55   3
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH55526 -  580.00 1123.13  3.3e-54   5
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAL20462 +  579.00 1116.76  7.4e-54   7
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAA54130 +  555.00 1074.68  1.6e-51   50
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAV88876 +  555.00 1073.25  2.0e-51   57
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAS43038 +  555.00 1072.49  2.2e-51   6
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAS12919 +  555.00 1072.49  2.2e-51   6
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAI29081 +  555.00 1072.05  2.3e-51   6
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAC86999 +  555.00 1070.66  2.7e-51   7
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:AAV19157 +  555.00 1070.60  2.8e-51   73
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAF63316 +  555.00 1070.60  2.8e-51   7
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAA54129 +  555.00 1070.36  2.8e-51   74
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAS47160 -  547.00 1063.77  6.6e-51   3
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAF17730 -  547.00 1063.77  6.6e-51
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH33168 +  542.00 1042.94  9.6e-50
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:ABK09584 -  535.00 1040.71  1.3e-49
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:ABK15919 +  529.00 1028.60  6.0e-49
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAA77841 +  516.00 1003.66  1.5e-47

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAZ52865

seq_documentation_block:
ID   AAZ52865 standard; cDNA; 1020 BP.
XX
AC   AAZ52865;
XX
DT   14-MAR-2000  (first entry)
XX
DE   Human prostate tumor cDNA library derived EST fragment #8.
XX
KW   Pancreas; tumor; EST; expressed sequence tag; human; cytostatic;
KW   treatment; ds.
XX
OS   Homo sapiens.
XX
PN   DE19820190-A1.
XX
PD   04-NOV-1999.
XX
PF   28-APR-1998;   98DE-1020190.
XX
PR   28-APR-1998;   98DE-1020190.
XX
PA   (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI   Rosenthal A,  Specht T,  Hinzmann B,  Schmitt A,  Pilarsky C,  Dahl E;
XX
DR   WPI; 1999-621386/54.
DR   P-PSDB; AAY73835, AAY73836, AAY73837.
XX
PT   New human nucleic acid sequences from pancreatic tumors, and related
PT   proteins  -
XX
PS   Claim 2; Page 188; 502pp; German.
XX
CC   This invention describes novel polypeptides and their encoding nucleic
CC   acids derived from human pancreatic tumor tissue which have cytostatic
CC   activity. The sequences are also useful in producing pharmaceutical
CC   compositions for treatment of pancreatic tumors. AAZ52858-Z53014
CC   represent expressed sequence tag (EST) fragments derived from a human
CC   pancreatic tumor cDNA library and which encode the proteins represented
CC   in AAY73814-Y74252.
XX
SQ   Sequence 1020 BP; 341 A; 214 C; 218 G; 247 T; 0 other;


alignment_scores:
            Quality: 1063.00              Length:    206
              Ratio:  5.160                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x AAZ52865   ..

 Align seg 1/1  to: AAZ52865  from: 1  to: 1020

       1 ArgLeuSerCysAlaGlyThrLeuSerGlySerGlyProHisProSerAr 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      57 AGACTCAGCTGTGCTGGCACACTCAGCGGCTCTGGACCGCATCCTAGCCG 106

      17 gArgLeuThrGlnGlyArgTrpValArgLysSerArgValAlaMetGluL 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     107 CCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAGAGTTGCCATGGAGA 156

      34 ysIleProValSerAlaPheLeuLeuLeuValAlaLeuSerTyrThrLeu 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
    157 AAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCTCCTACACTCTG 206

     51 AlaArgAspThrThrValLysProGlyAlaLysLysAspThrLysAspSe 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    207 GCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGACACAAAGGACTC 256

     67 rArgProLysLeuProGlnThrLeuSerArgGlyTrpGlyAspGlnLeuI 84
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    257 TCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGGGTGACCAACTCA 306

     84 leTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerLysThrSerAsn 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    307 TCTGGACTCAGACATATGAAGAAGCTCTATATAAATCCAAGACAAGCAAC 356

    101 LysProLeuMetIleIleHisHisLeuAspGluCysProHisSerGlnAl 117
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    357 AAACCCTTGATGATTATTCATCACTTGGATGAGTGCCCACACAGTCAAGC 406

    117 aLeuLysLysValPheAlaGluAsnLysGluIleGlnLysLeuAlaGluG 134
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    407 TTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCAGAAATTGGCAGAGC 456

    134 lnPheValLeuLeuAsnLeuValTyrGluThrThrAspLysHisLeuSer 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    457 AGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTGACAAACACCTTTCT 506

    151 ProAspGlyGlnTyrValProArgIleMetPheValAspProSerLeuTh 167
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    507 CCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTTGACCCATCTCTGAC 556

    167 rValArgAlaAspIleThrGlyArgTyrSerAsnArgLeuTyrAlaTyrG 184
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    557 AGTTAGAGCCGATATCACTGGAAGATATTCAAACCGTCTCTATGCTTACG 606

    184 luProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLysAlaLeuLys 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    607 AACCTGCAGATACAGCTCTGTTGCTTGACAACATGAAGAAAGCTCTCAAG 656

    201 LeuLeuLysThrGluLeu 206
        ||||||||||||||||||
    657 TTGCTGAAGACTGAATTG 674
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAC58380

seq_documentation_block:

ID AAC58380 standard; cDNA; 920 BP.
XX
AC AAC58380;
XX
DT 29-JAN-2001 (first entry)
XX
DE Human PRO1030 nucleotide sequence SEQ ID NO:39.
XX
KW Human; tumour; diagnosis; neoplastic disease; neoplastic cell growth;
KW proliferation; tumourigenesis; identification; cancer; cytostatic;
KW nootropic; neuroprotective; antiinflammatory; immunosuppressive;
KW immunostimulant; antiangiogenic; leukaemia; lymphoid malignancy;
KW neuronal disorder; glial disorder; astrocytal disorder; angiogenic;
KW hypothalamic disorder; glandular disorder; macrophagal disorder;
KW epithelial disorder; stromal disorder; blastocoelic disorder;
KW inflammatory disorder; immunologic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200053755-A2.
XX
PD 14-SEP-2000.
XX
PF 06-JAN-2000; 2000WO-US00376.
XX
PR 08-MAR-1999; 99WO-US05028.
PR 02-JUN-1999; 99WO-US12252.
PR 23-JUN-1999; 99US-0141037.
PR 07-JUL-1999; 99US-0143048.
PR 26-JUL-1999; 99US-0145698.
PR 30-NOV-1999; 99WO-US28313.
PR 20-DEC-1999; 99WO-US30911.
PR 05-JAN-2000; 2000WO-US00219.
XX
PA (GETH) GENENTECH INC.
XX
PI Ashkenazi AJ, Baker KP, Goddard A, Gurney AL, Hillan KJ, Roy MA;
PI Watanabe CK, Wood WI;
XX
DR WPI; 2000-572270/53.
DR P-PSDB; AAB24070.
XX
PT Thirty PRO polynucleotides encoding PRO polypeptides, useful in the
PT treatment, diagnosis and prevention of cancer -
XX
PS Claim 50; Fig 27; 286pp; English.
XX
CC The present invention describes an isolated antibody that binds to
CC one of the human PRO proteins designated PRO212, PRO290, PRO341, PRO535,
CC PRO619, PRO717, PRO809, PRO830, PRO848, PRO943, PRO1005, PRO1009,
CC PRO1025, PRO1030, PRO1097, PRO1107, PRO1111, PRO1153, PRO1182, PRO1184,
CC PRO1187, PRO1281, PRO23, PRO39, PRO834, PRO1317, PRO1710, PRO2094,
CC PRO2145 OR PRO2198. PRO antagonists can be used to inhibit tumour cell
CC growth. The PRO polypeptides and nucleotides are useful in the
CC treatment, diagnosis and prevention of cancer. The antibodies and other
CC anti-tumour compounds maybe used to treat various conditions, including
CC those characterised by overexpression and/or activation of the amplified
CC PRO genes. Exemplary conditions or disorders to be treated with such
CC antibodies and other compounds include benign or malignant tumours
CC (e.g., renal, liver, kidney, bladder, breast, gastric, ovarian,
CC colorectal, prostate, pancreatic, lung, vulva, thyroid, hepatic
CC carcinomas, sarcomas, glioblastomas, and various head and neck tumours),
CC leukaemias and lymphoid malignancies, other disorders such as neuronal,
CC glial, astrocytal, hypothalamic and other glandular, macrophagal,
CC epithelial, stromal and blastocoelic disorders, and inflammatory,
CC angiogenic and immunologic disorders. AAC58242 to AAC58366 represent PCR
CC primers and hybridisation probes used in the isolation of the human PRO
CC sequences. AAC58367 to AAC58396 and AAB24057 to AAB24089 represent human
CC PRO polynucleotide and protein sequences given in the exemplification of
CC the present invention.
XX
SQ Sequence 920 BP; 296 A; 198 C; 195 G; 231 T; 0 other;

alignment_scores:

```
          Quality: 1025.00          Length:    206
            Ratio:  5.025             Gaps:      0
 Percent Similarity:  99.029   Percent Identity: 98.544
```

alignment_block:

US-09-674-266A-181 x AAC58380 ..

Align seg 1/1 to: AAC58380 from: 1 to: 920

```
      1 ArgLeuSerCysAlaGlyThrLeuSerGlySerGlyProHisProSerAr 17
        |||||||||||||||||||||||    :::   ||||||||||||||||||
     56 AGACTCAGCTGTGCTGGCACACTCAGAAGCTT.GGACCGCATCCTAGCCG 104

     17 gArgLeuThrGlnGlyArgTrpValArgLysSerArgValAlaMetGluL 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    105 CCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAGAGTTGCCATGGAGA 154

     34 ysIleProValSerAlaPheLeuLeuLeuValAlaLeuSerTyrThrLeu 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    155 AAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCTCCTACACTCTG 204

     51 AlaArgAspThrThrValLysProGlyAlaLysLysAspThrLysAspSe 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    205 GCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGACACAAAGGACTC 254

     67 rArgProLysLeuProGlnThrLeuSerArgGlyTrpGlyAspGlnLeuI 84
```

```
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 255 TCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGGGTGACCAACTCA 304

  84 leTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerLysThrSerAsn 100
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 305 TCTGGACTCAGACATATGAAGAAGCTCTATATAAATCCAAGACAAGCAAC 354

 101 LysProLeuMetIleIleHisHisLeuAspGluCysProHisSerGlnAl 117
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 355 AAACCCTTGATGATTATTCATCACTTGGATGAGTGCCCACACAGTCAAGC 404

 117 aLeuLysLysValPheAlaGluAsnLysGluIleGlnLysLeuAlaGluG 134
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 405 TTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCAGAAATTGGCAGAGC 454

 134 lnPheValLeuLeuAsnLeuValTyrGluThrThrAspLysHisLeuSer 150
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 455 AGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTGACAAACACCTTTCT 504

 151 ProAspGlyGlnTyrValProArgIleMetPheValAspProSerLeuTh 167
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 505 CCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTTGACCCATCTCTGAC 554

 167 rValArgAlaAspIleThrGlyArgTyrSerAsnArgLeuTyrAlaTyrG 184
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 555 AGTTAGAGCCGATATCACTGGAAGATATTCAAATCGTCTCTATGCTTACG 604

 184 luProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLysAlaLeuLys 200
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 605 AACCTGCAGATACAGCTCTGTTGCTTGACAACATGAAGAAAGCTCTCAAG 654

 201 LeuLeuLysThrGluLeu 206
     ||||||||||||||||||
 655 TTGCTGAAGACTGAATTG 672
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAZ77486

seq_documentation_block:

```
ID   AAZ77486 standard; cDNA; 1018 BP.
XX
AC   AAZ77486;
XX
DT   10-APR-2000  (first entry)
XX
DE   Human ovarian tumor cDNA library derived EST fragment 37.
XX
KW   Expressed sequence tag; EST; human; ovarian tumor; anticancer;
KW   gene therapy; treatment; ss.
XX
OS   Homo sapiens.
XX
PN   DE19817557-A1.
XX
PD   21-OCT-1999.
XX
PF   09-APR-1998;   98DE-1017557.
XX
PR   09-APR-1998;   98DE-1017557.
XX
PA   (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI   Rosenthal A,  Specht T,  Hinzmann B,  Schmitt A,  Pilarsky C,  Dahl E;
XX
DR   WPI; 1999-591920/51.
DR   P-PSDB; AAY76590.
XX
PT   New nucleic acid sequences expressed in ovarian, and some other, cancer
PT   tissues, and derived polypeptides, for treatment of ovarian cancer and
PT   identification of therapeutic agents  -
XX
PS   Claim 3; Page 167; 310pp; German.
XX
CC   This invention describes novel nucleic acid (cDNA) sequences (A) which
CC   have anticancer activity and are highly expressed in ovarian tumor
CC   tissue (and some also in testis and breast cancer tissue). The products
CC   of the invention can be used for gene therapy. (A) are used (i) for
CC   recombinant expression of polypeptides (B) and (ii) to isolate complete
CC   genes. (B) are used (i) to identify agents suitable for treatment of
CC   ovarian cancer; (ii) directly for treating this form of cancer
CC   (including expression from gene therapy vectors) and (iii) for generation
CC   of specific antibodies. (A) are identified by assembling ESTs (expressed
CC   sequence tags) from a particular tissue type before comparison of
CC   expression patterns. This allows a significantly longer fragment of the
CC   gene to be revealed, so should reduce the number of failures associated
CC   with the fact that ESTs from different libraries may represent different
CC   parts of the same unknown gene, distorting the estimated frequency of
CC   occurrence in a particular tissue. AAZ77450-Z77572 represent the human
CC   ovarian tumor cDNA library derived EST fragments described in the method
CC   of the invention and encode the protein fragments represented in
CC   AAY76505-Y76638.
XX
SQ   Sequence 1018 BP; 343 A; 212 C; 216 G; 247 T; 0 other;
```

alignment_scores:

```
            Quality: 1018.50              Length:    207
              Ratio:  5.017                 Gaps:      2
 Percent Similarity:  98.068   Percent Identity: 98.068
```

alignment_block:

US-09-674-266A-181 x AAZ77486 ..

Align seg 1/1 to: AAZ77486 from: 1 to: 1018

```
       1 ArgLeuSerCysAlaGlyThrLeuSerGlySer.GlyProHisProSerA 17
         |||||||||||||||||||||||      |||    ||||||||||||
      57 AGACTCAGCTGTGCTGGCACACTCAGA...AGCTTGACCGCATCCTAGCC 103

      17 rgArgLeuThrGlnGlyArgTrpValArgLysSerArgValAlaMetGlu 33
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     104 GCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAGAGTTGCCATGGAG 153

      34 LysIleProValSerAlaPheLeuLeuLeuValAlaLeuSerTyrThrLe 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     154 AAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCTCCTACACTCT 203

      50 uAlaArgAspThrThrValLysProGlyAlaLysLysAspThrLysAspS 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     204 GGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGACACAAAGGACT 253

      67 erArgProLysLeuProGlnThrLeuSerArgGlyTrpGlyAspGlnLeu 83
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     254 CTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGGGTGACCAACTC 303

      84 IleTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerLysThrSerAs 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     304 ATCTGGACTCAGACATATGAAGAAGCTCTATATAAATCCAAGACAAGCAA 353

     100 nLysProLeuMetIleIleHisHisLeuAspGluCysProHisSerGlnA 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     354 CAAACCCTTGATGATTATTCATCACTTGGATGAGTGCCCACACAGTCAAG 403

     117 laLeuLysLysValPheAlaGluAsnLysGluIleGlnLysLeuAlaGlu 133
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     404 CTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCAGAAATTGGCAGAG 453

     134 GlnPheValLeuLeuAsnLeuValTyrGluThrThrAspLysHisLeuSe 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     454 CAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTGACAAACACCTTTC 503

     150 rProAspGlyGlnTyrValProArgIleMetPheValAspProSerLeuT 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     504 TCCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTTGACCCATCTCTGA 553

     167 hrValArgAlaAspIleThrGlyArgTyrSerAsnArgLeuTyrAlaTyr 183
         |||||||||||||||||||||||||||||||||||||||||||||||||
```

```
     554 CAGTTAGAGCCGATATCACTGGAAGATATTCAAACCGTCTCTATGCTTAC 603

     184 GluProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLysAlaLeuLy 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     604 GAACCTGCAGATACAGCTCTGTTGCTTGACAACATGAAGAAAGCTCTCAA 653

     200 sLeuLeuLysThrGluLeu 206
         ||||||||||||||||||
     654 GTTGCTGAAGACTGAATTG 672

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:AAV29047

seq_documentation_block:
ID   AAV29047 standard; cDNA; 866 BP.
XX
AC   AAV29047;
XX
DT   21-AUG-1998  (first entry)
XX
DE   Human protein comprising secretory signal nucleotide sequence 9.
XX
KW   Human protein; secretory signal; nutritional source; cytokine;
KW   immunity;haematopoiesis; activin; inhibin; tumour; chemotactic;
KW   chemokinetic; thrombolytic; anti-inflammatory; inhibition;
KW   stomach cancer cell; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             73..600
FT                   /*tag= a
FT                   /product= "human protein comprising secretory
FT                            signal"
XX
PN   WO9811217-A2.
XX
PD   19-MAR-1998.
XX
PF   12-SEP-1997;   97WO-JP03239.
XX
PR   13-SEP-1996;   96JP-0243060.
XX
PA   (PROT-) PROTEGENE INC.
PA   (SAGA ) SAGAMI CHEM RES CENTRE.
XX
PI   Kato S,  Kobayashi M,  Sekine S,  Yamaguchi T;
XX
DR   WPI; 1998-207380/18.
DR   P-PSDB; AAW37872.
XX
PT   Human proteins with secretory signal sequences - used to treat
PT   immune deficiencies, infections, tumours, and haematopoietic
PT   disorders, etc.
XX
PS   Claim 4; Pages 114-116; 131pp; English.
XX
CC   This is the nucleotide sequence of a novel human protein comprising
CC   a secretory signal isolated from stomach cancer cells.  Its proteins
CC   can be used as nutritional sources or supplements.  The proteins may
CC   also have cytokine functions, immune modulating functions,
CC   haematopoiesis regulating activity, activin/inhibin regulating
CC   activity, chemotactic/chemokinetic activity, haemostatic and
CC   thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC   activity, tumour inhibition activity.
XX
SQ   Sequence 866 BP; 275 A; 189 C; 178 G; 224 T; 0 other;


alignment_scores:
           Quality: 1008.00            Length:    195
             Ratio:  5.169              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x AAV29047    ..

 Align seg 1/1  to: AAV29047  from: 1  to: 866

      12 GlyProHisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSe 28
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      13 GGACCGCATCCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATC 62

      28 rArgValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValA 45
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      63 CAGAGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGG 112

      45 laLeuSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLys 61
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     113 CCCTCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAA 162

      62 LysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGl 78
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     163 AAGGACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGG 212

      78 yTrpGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrL 95
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     213 TTGGGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATA 262

      95 ysSerLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGlu 111
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     263 AATCCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAG 312

     112 CysProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIl 128
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     313 TGCCCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAAT 362

     128 eGlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrT 145
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     363 CCAGAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAA 412

     145 hrAspLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPhe 161
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     413 CTGACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTT 462

     162 ValAspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAs 178
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     463 GTTGACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAA 512

     178 nArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnM 195
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     513 CCGTCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACA 562

     195 etLysLysAlaLeuLysLeuLeuLysThrGluLeu 206
         ||||||||||||||||||||||||||||||||||
     563 TGAAGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 597

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH35026

seq_documentation_block:
ID   AAH35026 standard; cDNA; 943 BP.
XX
AC   AAH35026;
XX
DT   03-SEP-2001  (first entry)
XX
DE   Human colon cancer antigen encoding cDNA SEQ ID NO:2108.
XX
KW   Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW   colorectal carcinoma; ss.
XX
OS   Homo sapiens.
XX
PN   WO200122920-A2.
XX
PD   05-APR-2001.
XX
```

```
PF   28-SEP-2000; 2000WO-US26524.
XX
PR   29-SEP-1999;   99US-0157137.
PR   03-NOV-1999;   99US-0163280.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM, Barash SC,  Birse CE,  Rosen CA;
XX
DR   WPI; 2001-235357/24.
DR   P-PSDB; AAG75621.
XX
PT   Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT   useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS   Claim 1; Page 3593; 9803pp; English.
XX
CC   AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC   cancer-associated nucleic acid molecules (N) and proteins (P), where
CC   the proteins are collectively known as colon cancer antigens. The colon
CC   cancer antigens have cytostatic activity and can be used in gene
CC   therapy and vaccine production. N and P may be used in the prevention,
CC   diagnosis and treatment of diseases associated with inappropriate P
CC   expression. For example, N and P may be used to treat disorders
CC   associated with decreased expression by rectifying mutations or deletions
CC   in a patient's genome that affect the activity of P by expressing
CC   inactive proteins or to supplement the patients own production of P.
CC   Additionally, N may be used to produce the colon cancer-associated Ps,
CC   by inserting the nucleic acids into a host cell and culturing the cell
CC   to express the proteins. N and P can be used in the prevention, diagnosis
CC   and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC   and AAB77789 represent sequences used in the exemplification of the
CC   present invention.
CC   N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC   missing at time of publication, meaning no sequences are present for
CC   SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ   Sequence 943 BP; 311 A; 199 C; 199 G; 229 T; 5 other;


alignment_scores:
           Quality: 1002.00              Length:    194
             Ratio:  5.165                Gaps:      0
 Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x AAH35026   ..

 Align seg 1/1  to: AAH35026  from: 1  to: 943

      13 ProHisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerAr 29
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      26 CCGCATCCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAG 75

      29 gValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAlaL 46
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      76 AGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCC 125

      46 euSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLys 62
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     126 TCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAG 175

      63 AspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTr 79
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     176 GACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTG 225

      79 pGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysS 96
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     226 GGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAAT 275

      96 erLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCys 112
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     276 CCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGC 325

     113 ProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGl 129
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     326 CCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCA 375

     129 nLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrA 146
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     376 GAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTG 425

     146 spLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheVal 162
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     426 ACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTT 475

     163 AspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnAr 179
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     476 GACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAAYCG 525

     179 gLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetL 196
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     526 TCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACATGA 575

     196 ysLysAlaLeuLysLeuLeuLysThrGluLeu 206
         |||||||||||||||||||||||||||||||
     576 AGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 607

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:AAV19155

seq_documentation_block:
ID   AAV19155 standard; cDNA; 875 BP.
XX
AC   AAV19155;
XX
DT   28-AUG-1998  (first entry)
XX
DE   Human XAG growth factor huXAG-1 cDNA.
XX
KW   HuXAG-1; XAG; growth factor; colon cancer-specific gene;
KW   tumour marker; breast disease; liver disease; lung disease;
KW   emphysema; wound healing; diagnosis; therapy; human; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             71..598
FT                   /*tag=  a
FT   sig_peptide     71..130
FT                   /*tag=  b
FT   mat_peptide     131..595
FT                   /*tag=  c
XX
PN   WO9807749-A1.
XX
PD   26-FEB-1998.
XX
PF   22-AUG-1997;   97WO-US14139.
XX
PR   23-AUG-1996;   96WO-US13766.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Dillon PJ,  Ebner R,  Endress GA,  Yu G;
XX
DR   WPI; 1998-169093/15.
DR   P-PSDB; AAW37844.
XX
PT   New isolated human XAG growth factor(s) - used to develop products
PT   for treating e.g. liver, lung or breast diseases or
PT   hyperproliferative disorders, e.g. cancer.
XX
PS   Claim 5; Fig 1; 141pp; English.
XX
CC   This cDNA clone codes for huXAG-1 (see AAW37844), a member of a novel
CC   family of human growth factors also including huXAG-2 (see AAW37845)
```

```
CC   and huXAG-3 (see AAW37846). These proteins share homology with the
CC   XAG protein of Xenopus laevis, which is involved in embryogenesis
CC   and is expressed in adult tissue. Expression of huXAG-1 has been
CC   discovered in colon cancer tissue, with no corresponding expression
CC   in healthy colon tissue. The huXAG-1 gene, also designated
CC   colon cancer-specific gene (CCSG), provides a molecular marker for
CC   colon cancer. huXAG-1 cDNA was isolated from a cDNA library
CC   derived from human colon cancer tissue. Vectors, host cells,
CC   antibodies, and screening methods for identifying agonists and
CC   antagonists of huXAG-1 are provided. HuXAG polypeptides are growth
CC   factors and can be used to stimulate proliferation of cells. They
CC   can be used to stimulate the proliferation and differentiation of
CC   hepatocytes to alleviate or treat liver diseases and pathologies
CC   such as fulminant liver failure caused by cirrhosis, liver damage
CC   caused by viral hepatitis and toxic substances. They can also be
CC   used to stimulate or promote liver regeneration, e.g. after
CC   surgery. They can also be used to prevent and heal damage to the
CC   lungs caused by various pathological states. They can be used to
CC   stimulate proliferation and differentiation and promote the repair
CC   of alveoli and bronchiolar epithelium to prevent, attenuate, or
CC   treat acute or chronic lung damage, e.g. emphysema, which results
CC   in the progressive loss of alveoli, and inhalation injuries, e.g.
CC   resulting from smoke inhalation and burns, that cause necrosis of
CC   the bronchiolar epithelium and alveoli. They can also be used to
CC   stimulate the proliferation and differentiation of breast tissue
CC   and could therefore be used to promote healing of breast tissue
CC   injury due to surgery, trauma or cancer. Antagonists can be used
CC   to treat hyperproliferative disorders, including cancer, in
CC   particular hepatocellular carcinoma, osteoclastoma, breast cancer,
CC   or colon cancer. The products can also be used for detection and
CC   diagnosis.
XX
SQ   Sequence 875 BP; 283 A; 189 C; 180 G; 223 T; 0 other;


alignment_scores:
           Quality:  991.00              Length:    194
             Ratio:  5.135                 Gaps:      0
 Percent Similarity:  99.485   Percent Identity: 99.485

alignment_block:
 US-09-674-266A-181 x AAV19155   ..

 Align seg 1/1  to: AAV19155  from: 1  to: 875

      13 ProHisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerAr 29
         |||   ||||||||||||||||||||||||||||||||||||||||||||
      14 CCGATTCCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAG 63

      29 gValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAlaL 46
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      64 AGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCC 113

      46 euSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLys 62
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     114 TCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAG 163

      63 AspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTr 79
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     164 GACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTG 213

      79 pGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysS 96
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     214 GGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAAT 263

      96 erLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCys 112
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     264 CCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGC 313

     113 ProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGl 129
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     314 CCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCA 363

     129 nLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrA 146
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     364 GAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTG 413

     146 spLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheVal 162
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     414 ACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTT 463

     163 AspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnAr 179
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     464 GACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAATCG 513

     179 gLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetL 196
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     514 TCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACATGA 563

     196 ysLysAlaLeuLysLeuLeuLysThrGluLeu 206
         |||||||||||||||||||||||||||||||
     564 AGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 595

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAF63314

seq_documentation_block:
ID   AAF63314 standard; cDNA; 875 BP.
XX
AC   AAF63314;
XX
DT   04-MAY-2001  (first entry)
XX
DE   Human huXAG-1/CCSG colon cancer specific gene cDNA.
XX
KW   Human; growth factor; huXAG-1; colon cancer specific gene; CCSG;
KW   cell proliferation; liver disease; fulminant liver failure; cirrhosis;
KW   hepatitis; cancer; colon cancer; colorectal carcinoma; ss.
XX
OS   Homo sapiens.
XX
PN   US6171816-B1.
XX
PD   09-JAN-2001.
XX
PF   22-AUG-1997; 97US-0916576.
XX
PR   23-AUG-1996; 96US-0024347.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Yu G, Dillon PJ, Ebner R, Endress GA;
XX
DR   WPI; 2001-136872/14.
DR   P-PSDB; AAB72203.
XX
PT   Novel human growth factor polypeptide useful for diagnosing and
PT   treating colon cancer and liver diseases, to prevent and heal damage to
PT   the lungs and for identifying modulators of therapeutic use -
XX
PS   Claim 45; Fig 1; 59pp; English.
XX
CC   This invention relates to a human growth factor polypeptide huXAG-1 also
CC   known as a colon cancer specific gene (CCSG). HuXAG-1 stimulates cell
CC   proliferation as a growth factor. The HuXAG-1 protein is useful for
CC   identifying compounds capable of enhancing or inhibiting cellular
CC   response induced by huXAG-1. The protein is also useful for stimulating
CC   proliferation of cells e.g. colon, breast, liver and lung cells, and
CC   hepatocytes. It is useful for alleviating or treating liver diseases and
CC   pathologies such as fulminant liver failure caused by cirrhosis, liver
CC   damage caused by viral hepatitis and toxic substances, for preventing and
CC   treating damage to the lungs caused by various pathological states and
CC   for promoting healing of breast tissue injury due to surgery, trauma or
CC   cancer. HuXAG-1 and the identified antagonist are useful for treating
CC   cancer, in particular colon cancer. Detecting altered levels of huXAG-1
CC   and its polynucleotides are useful for diagnosing or detecting cancer in
CC   mammals. The gene encoding huXAG-1 is useful for monitoring human
```

```
CC   colorectal carcinoma. huXAG-1 nucleic acid molecules are also useful for
CC   chromosome identification. The present sequence represents cDNA encoding
CC   huXAG-1.
XX
SQ   Sequence 875 BP; 283 A; 189 C; 180 G; 223 T; 0 other;


alignment_scores:
            Quality:  991.00              Length:    194
              Ratio:  5.135                 Gaps:      0
 Percent Similarity:  99.485   Percent Identity: 99.485

alignment_block:
 US-09-674-266A-181 x AAF63314   ..

 Align seg 1/1  to: AAF63314  from: 1  to: 875

      13 ProHisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerAr 29
         |||   |||||||||||||||||||||||||||||||||||||||||||||
      14 CCGATTCCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAG 63

      29 gValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAlaL 46
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      64 AGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCC 113

      46 euSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLys 62
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     114 TCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAG 163

      63 AspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTr 79
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     164 GACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTG 213

      79 pGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysS 96
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     214 GGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAAT 263

      96 erLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCys 112
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     264 CCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGC 313

     113 ProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGl 129
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     314 CCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCA 363

     129 nLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrA 146
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     364 GAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTG 413

     146 spLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheVal 162
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     414 ACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTT 463

     163 AspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnAr 179
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     464 GACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAATCG 513

     179 gLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetL 196
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     514 TCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACATGA 563

     196 ysLysAlaLeuLysLeuLeuLysThrGluLeu 206
         |||||||||||||||||||||||||||||||
     564 AGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 595

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:AAV59320

seq_documentation_block:
ID   AAV59320 standard; cDNA; 881 BP.
XX
AC   AAV59320;
XX
DT   21-DEC-1998  (first entry)
XX
DE   Nucleotide sequence encoding zsig10 polypeptide.
XX
KW   ss; human; mucous-mediated function; adhesion; tumour metastasis;
KW   bacterial colonisation; microbial infection; AIDS; cystic fibrosis;
KW   chronic obstructive pulmonary disease; asthma; Crohn's disease;
KW   sinonasal inflammatory disease; inflammatory bowel disease; bronchitis.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             63..590
FT                   /*tag= a
FT                   /product= "zsig10"
XX
PN   WO9841627-A1.
XX
PD   24-SEP-1998.
XX
PF   18-MAR-1998;   98WO-US05251.
XX
PR   19-MAR-1997;   97US-0039631.
XX
PA   (ZYMO ) ZYMOGENETICS INC.
XX
PI   Sheppard PO;
XX
DR   WPI; 1998-531566/45.
DR   P-PSDB; AAW77365.
XX
PT   New isolated mucous-associated polypeptide, zsig10 - used to develop
PT   products for treating e.g. tumour metastasis, microbial infections,
PT   cystic fibrosis, asthma, bronchitis or inflammatory bowel disease
XX
PS   Claim 37; Page 80-81; 109pp; English.
XX
CC   The human polypeptide zsig10 is involved in mucous-mediated functions
CC   such as adhesion.  The products of the invention can be used in the study
CC   and treatment of e.g. tumour metastasis, bacterial colonisation,
CC   susceptibility to and persistence of infection, microbial infections,
CC   AIDS, cystic fibrosis, chronic obstructive pulmonary disease, asthma,
CC   sinonasal inflammatory disease, inflammatory bowel disease, bronchitis,
CC   or Crohn's disease.  The products can also be used for detection,
CC   diagnosis and drug screening.
XX
SQ   Sequence 881 BP; 293 A; 185 C; 180 G; 223 T; 0 other;


alignment_scores:
            Quality:  987.00              Length:    193
              Ratio:  5.141                 Gaps:      0
 Percent Similarity:  99.482   Percent Identity: 99.482

alignment_block:
 US-09-674-266A-181 x AAV59320   ..

 Align seg 1/1  to: AAV59320  from: 1  to: 881

      14 HisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerArgVa 30
         |||   |||||||||||||||||||||||||||||||||||||||||||||
       9 CACGAGAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAGAGT 58

      30 lAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAlaLeuS 47
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      59 TGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCT 108

      47 erTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLysAsp 63
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     109 CCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGAC 158

      64 ThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTrpGl 80
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     159 ACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGGG 208
```

```
 80 yAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerL 97
    ||||||||||||||||||||||||||||||||||||||||||||||||||
209 TGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAATCCA 258

 97 ysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCysPro 113
    ||||||||||||||||||||||||||||||||||||||||||||||||||
259 AGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGCCCA 308

114 HisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGlnLy 130
    ||||||||||||||||||||||||||||||||||||||||||||||||||
309 CACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCAGAA 358

130 sLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrAspL 147
    ||||||||||||||||||||||||||||||||||||||||||||||||||
359 ATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTGACA 408

147 ysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheValAsp 163
    ||||||||||||||||||||||||||||||||||||||||||||||||||
409 AACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTTGAC 458

164 ProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnArgLe 180
    ||||||||||||||||||||||||||||||||||||||||||||||||||
459 CCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAATCGTCT 508

180 uTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetLysL 197
    ||||||||||||||||||||||||||||||||||||||||||||||||||
509 CTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACATGAAGA 558

197 ysAlaLeuLysLeuLeuLysThrGluLeu 206
    |||||||||||||||||||||||||||||
559 AAGCTCTCAAGTTGCTGAAGACTGAATTG 587
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAF44884

seq_documentation_block:

```
ID   AAF44884 standard; cDNA; 778 BP.
XX
AC   AAF44884;
XX
DT   28-MAR-2001  (first entry)
XX
DE   Human breast cancer related protein coding sequence SEQ ID NO: 40.
XX
KW   Human; breast cancer; diagnosis; therapy; vaccine; ss.
XX
OS   Homo sapiens.
XX
PN   WO200078960-A2.
XX
PD   28-DEC-2000.
XX
PF   23-JUN-2000; 2000WO-US17536.
XX
PR   23-JUN-1999;   99US-0140903.
PR   12-OCT-1999;   99US-0158980.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Yuqiu J,  Mitcham JL;
XX
DR   WPI; 2001-041426/05.
XX
PT   New polynucleotides encoding breast tumor specific proteins, useful for
PT   prevention, treatment and diagnosis of breast cancer -
XX
PS   Claim 25; Page 132; 165pp; English.
XX
CC   The present invention provides the coding sequences for a number of
CC   breast cancer related proteins. These can be used in vaccinations
CC   against, diagnosis of and treatment of cancer, particularly breast
CC   cancer.
XX
```

SQ Sequence 778 BP; 245 A; 175 C; 158 G; 193 T; 7 other;

alignment_scores:

```
           Quality:  967.00            Length:    194
             Ratio:  5.036               Gaps:      0
 Percent Similarity:  98.969   Percent Identity: 98.969
```

alignment_block:

US-09-674-266A-181 x AAF44884 ..

Align seg 1/1 to: AAF44884 from: 1 to: 778

```
 13 ProHisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerAr 29
    ||||||||||||||||||||||||||||||||||||||||||||||||||
  2 CCGCATCCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAG 51

 29 gValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAlaL 46
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 52 AGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCC 101

 46 euSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLys 62
    ||||||||||||||||||||   ||||||||||||||||||||||||||
102 TCTCCTACACTCTGGCCAGASATACCACAGTCAAACCTGGAGCCAAAAAG 151

 63 AspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTr 79
    ||||||||||||||||||||||||||||||||||||||||||||||||||
152 GACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTG 201

 79 pGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysS 96
    ||||||||||||||||||||||||||||||||||||||||||||||||||
202 GGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAAT 251

 96 erLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCys 112
    ||||||||||||||||||||||||||||||||||||||||||||||||||
252 CCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGC 301

113 ProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGl 129
    ||||||||||||||||||||||||||||||||||||||||||||||||||
302 CCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCA 351

129 nLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrA 146
    ||||||||||||||||||||||||||||||||||||||||||||||||||
352 GAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTG 401

146 spLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheVal 162
    ||||||||||||||||||||||||||||||||||||||||||||||||||
402 ACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTT 451

163 AspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnAr 179
    ||||||||||||||||||||||||||||||||||||||||||||||||||
452 GACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAATCG 501

179 gLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetL 196
    ||||||||||||||||||||||   ||||||||||||||||||||||||
502 TCTCTATGCTTACGAACCTGCAGATCA.GCTCTGTTGCTTGACAACATGA 550

196 ysLysAlaLeuLysLeuLeuLysThrGluLeu 206
    |||||||||||||||||||||||||||||||
551 AGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 582
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAZ40803

seq_documentation_block:

```
ID   AAZ40803 standard; DNA; 1689 BP.
XX
AC   AAZ40803;
XX
DT   18-JAN-2000  (first entry)
XX
DE   Secreted protein EST coding sequence 108-008-5-0-A6-FL.
XX
KW   Secreted protein; fingerprint identification technique;
```

```
KW   chromosome mapping; human; hereditary disease; diagnosis; cancer;
KW   hyperlipidaemia; cardiovascular; neurodegenerative disorder; therapy;
KW   autoimmune disease; rheumatic disease; embryogenic disorder; myopathy;
KW   renal injury; amino aciduria; hypoglycaemia; male rat infertility;
KW   hypertension; ss.
XX
OS   Homo sapiens.
XX
PN   WO9940189-A2.
XX
PD   12-AUG-1999.
XX
PF   09-FEB-1999;   99WO-IB00282.
XX
PR   09-FEB-1998;   98US-0074121.
PR   13-APR-1998;   98US-0081563.
PR   10-AUG-1998;   98US-0096116.
PR   04-SEP-1998;   98US-0099273.
XX
PA   (GEST ) GENSET.
XX
PI   Bougueleret L,  Duclert A,  Dumas Milne Edwards J;
XX
DR   WPI; 1999-600966/51.
DR   P-PSDB; AAY59675.
XX
PT   Extended cDNAs useful for expressing secreted proteins and to obtain
PT   specific antibodies -
XX
PS   Claim 1; Page 168-169; 244pp; English.
XX
CC   This sequence encodes a human secreted protein of the invention. The
CC   extended cDNAs (or genomic DNAs obtainable from them) may be used to
CC   prepare PCR primers and probes. These are useful for forensic matching or
CC   positive identification by DNA sequencing. They may also be used in
CC   alternative fingerprint identification techniques. Antibodies against the
CC   proteins encoded by the extended cDNAs are useful in identification of
CC   tissue types or cell species, as well as identifying tissue specific
CC   soluble proteins. The sequences can be used for chromosome mapping and
CC   identification of genes associated with hereditary diseases or drug
CC   response. signal sequences from the cDNAs can be used in construction of
CC   secretion vectors. Other sequences derived from the extended cDNAs can be
CC   used to clone upstream genomic DNA sequences including promoters. This is
CC   in turn useful for identifying proteins that interact with promoter
CC   sequences. Some of the proteins may be useful in diagnosing and treating
CC   several disorders including, but not limited to: cancer, hyperlipidaemia,
CC   cardiovascular and neurodegenerative disorders, autoimmune diseases, and
CC   rheumatic diseases, embryogenic disorders, hypertension, renal injury,
CC   amino acidurias, hypoglycaemia, male rat infertility and myopathies.
XX
SQ   Sequence 1689 BP; 552 A; 350 C; 335 G; 452 T; 0 other;


alignment_scores:
            Quality:  961.00              Length:    195
              Ratio:  5.085                 Gaps:      1
 Percent Similarity:  96.923   Percent Identity: 96.923

alignment_block:
 US-09-674-266A-181 x AAZ40803   ..

 Align seg 1/1  to: AAZ40803  from: 1  to: 1689

      12 GlyProHisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSe 28
         ||||||||||||||||||||||||||||||||
       9 GGACCGCATCCTAGCCGCCGACTCACACAAGGC..................... 41

      28 rArgValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValA 45
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      42 .AGAGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGG 90

      45 laLeuSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLys 61
         ||||||||||||||||||||||||||||||||||||||||||||||||||||

      91 CCCTCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAA 140

      62 LysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGl 78
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     141 AAGGACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGG 190

      78 yTrpGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrL 95
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     191 TTGGGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATA 240

      95 ysSerLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGlu 111
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     241 AATCCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAG 290

     112 CysProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIl 128
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     291 TGCCCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAAT 340

     128 eGlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrT 145
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     341 CCAGAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAA 390

     145 hrAspLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPhe 161
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     391 CTGACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTT 440

     162 ValAspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAs 178
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     441 GTTGACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAA 490

     178 nArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnM 195
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     491 TCGTCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACA 540

     195 etLysLysAlaLeuLysLeuLeuLysThrGluLeu 206
         ||||||||||||||||||||||||||||||||||
     541 TGAAGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 575

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAZ40846

seq_documentation_block:
ID   AAZ40846 standard; DNA; 940 BP.
XX
AC   AAZ40846;
XX
DT   18-JAN-2000  (first entry)
XX
DE   Secreted protein EST coding sequence 78-21-1-B7-FL1.
XX
KW   Secreted protein; fingerprint identification technique;
KW   chromosome mapping; human; hereditary disease; diagnosis; cancer;
KW   hyperlipidaemia; cardiovascular; neurodegenerative disorder; therapy;
KW   autoimmune disease; rheumatic disease; embryogenic disorder; myopathy;
KW   renal injury; amino aciduria; hypoglycaemia; male rat infertility;
KW   hypertension; ss.
XX
OS   Homo sapiens.
XX
PN   WO9940189-A2.
XX
PD   12-AUG-1999.
XX
PF   09-FEB-1999;   99WO-IB00282.
XX
PR   09-FEB-1998;   98US-0074121.
PR   13-APR-1998;   98US-0081563.
PR   10-AUG-1998;   98US-0096116.
PR   04-SEP-1998;   98US-0099273.
XX
PA   (GEST ) GENSET.
XX
PI   Bougueleret L,  Duclert A,  Dumas Milne Edwards J;
XX
```

```
DR   WPI; 1999-600966/51.
DR   P-PSDB; AAY59718.
XX
PT   Extended cDNAs useful for expressing secreted proteins and to obtain
PT   specific antibodies -
XX
PS   Claim 1; Page 226; 244pp; English.
XX
CC   This sequence encodes a human secreted protein of the invention. The
CC   extended cDNAs (or genomic DNAs obtainable from them) may be used to
CC   prepare PCR primers and probes. These are useful for forensic matching or
CC   positive identification by DNA sequencing. They may also be used in
CC   alternative fingerprint identification techniques. Antibodies against the
CC   proteins encoded by the extended cDNAs are useful in identification of
CC   tissue types or cell species, as well as identifying tissue specific
CC   soluble proteins. The sequences can be used for chromosome mapping and
CC   identification of genes associated with hereditary diseases or drug
CC   response. signal sequences from the cDNAs can be used in construction of
CC   secretion vectors. Other sequences derived from the extended cDNAs can be
CC   used to clone upstream genomic DNA sequences including promoters. This is
CC   in turn useful for identifying proteins that interact with promoter
CC   sequences. Some of the proteins may be useful in diagnosing and treating
CC   several disorders including, but not limited to: cancer, hyperlipidaemia,
CC   cardiovascular and neurodegenerative disorders, autoimmune diseases, and
CC   rheumatic diseases, embryogenic disorders, hypertension, renal injury,
CC   amino acidurias, hypoglycaemia, male rat infertility and myopathies.
XX
SQ   Sequence 940 BP; 304 A; 202 C; 187 G; 233 T; 14 other;


alignment_scores:
            Quality:  914.00              Length:    207
              Ratio:  4.811                Gaps:      2
 Percent Similarity:  91.787   Percent Identity: 90.821

alignment_block:
 US-09-674-266A-181 x AAZ40846   ..

 Align seg 1/1  to: AAZ40846  from: 1  to: 940

       1 ArgLeuSerCysAlaGlyThrLeuSerGlySerGlyProHisProSerAr 17
         ||||||||||||||||||||||||   :::  ||||||||||||||||||
     103 AGACTCAGCTGTGCTGGCACACTCAGAAGCTT.GGACCGCATCCTAGCCG 151

      17 gArgLeuThrGlnGlyArgTrpValArgLysSerArgValAlaMetGluL 34
         ||||||||||||||||                ||||||||||||||||
     152 CCGACTCACACAAGGC................AGAGTTGCCATGGAAA 183

      34 ysIleProValSerAlaPheLeuLeuLeuValAlaLeuSerTyrThrLeu 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     184 AAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCTCCTACACTCTG 233

      51 AlaArgAspThrThrValLysProGlyAlaLysLysAspThrLysAspSe 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     234 GCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGACACAAAGGACTC 283

      67 rArgProLysLeuProGlnThrLeuSerArgGlyTrpGlyAspGlnLeuI 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     284 TCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGGGTGACCAACTCA 333

      84 leTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerLysThrSerAsn 100
         ||||||||||||||||||   :::|||   ||||||||||||||||||
     334 TCTGGACACARACATATGAARAARCTCTATWTAAATCCAARACAAGCAAC 383

     101 LysProLeuMetIleIleHisHisLeuAspGluCysProHisSerGlnAl 117
         ||||||||||||||||||||||||||||   ||||||||||||||||||
     384 AAACCCTTGATGATTATTCATCACTTGGATGADTGCCCACACAGTCAAGC 433

     117 aLeuLysLysValPheAlaGluAsnLysGluIleGlnLysLeuAlaGluG 134
         ||||||||||   ||||||||||||||   ||||||||||||||||   |
     434 TTTAAAAAAAKTGTTTGCTGAAAATAAARAAATCCAGAAATTGGCARANC 483

     134 lnPheValLeuLeuAsnLeuValTyrGluThrThrAspLysHisLeuSer 150
         ||||||||   ||||||||||||||||||||||||||||||||||||||
     484 AGTTTGTCCYCCTCAATCTGGTTTATGAAACAACTGACAAACACCTTTCT 533

     151 ProAspGlyGlnTyrVal.ProArgIleMetPheValAspProSerLeuT 167
         ||||||||||||||    ||||||||||||||||||||||||||||||
     534 CCTGATGGCCAATATKTCCCCCMGGATTATGTTTGTTGACCCATCTCTGA 583

     167 hrValArgAlaAspIleThrGlyArgTyrSerAsnArgLeuTyrAlaTyr 183
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     584 CAGTTAGAGCCGATATCACTGGAAGATATTCAAAYCGTCTCTATGCTTAC 633

     184 GluProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLysAlaLeuLy 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     634 GAACCTGCAGATACAGCTCTGTTGCTTGACAACATGAAGAAAGCTCTCAA 683

     200 sLeuLeuLysThrGluLeu 206
         ||||||||||||||||||
     684 GTTGCTGAAGACTGAATTG 702

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAS13480

seq_documentation_block:
ID   AAS13480 standard; DNA; 543 BP.
XX
AC   AAS13480;
XX
DT   19-DEC-2001  (first entry)
XX
DE   DNA encoding breast cancer cell membrane protein 7 (BCMP 7).
XX
KW   Breast cancer cell membrane protein 7; BCMP 7; breast cancer;
KW   metastasis; cytostatic.; gene therapy; antibody; antisense;
KW   vaccine; chromosome 7p21.3; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             11..538
FT                   /*tag= a
FT                   /product= "BCMP 7"
FT                   /note= "Breast cancer cell membrane protein 7"
FT   sig_peptide     11..70
FT                   /*tag= b
FT   mat_peptide     71..535
FT                   /*tag= c
FT                   /note= "Mature BCMP 7"
XX
PN   WO200163290-A1.
XX
PD   30-AUG-2001.
XX
PF   21-FEB-2001; 2001WO-GB00734.
XX
PR   25-FEB-2000; 2000GB-0004576.
XX
PA   (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI   Boyd RS,  Stamps AC,  Terrett JA,  Tyson KL;
XX
DR   WPI; 2001-570651/64.
DR   P-PSDB; AAU08804.
XX
PT   Diagnosing, preventing and treating breast cancer using a breast cancer
PT   cell membrane protein BCMP 7 -
XX
PS   Claim 6; Fig 1; 62pp; English.
XX
CC   The invention describes the novel use of a protein found in breast
CC   cancer cell membranes (BCMP 7) for diagnosing, preventing and treating
CC   breast cancers. The peptide has cytostatic action and potential uses in
CC   gene therapy and in vaccines. The polypeptide, antisense nucleic acids,
CC   or fusion proteins comprising and Green Fluorescent protein or the DsRed
CC   Fluorescent protein, antibodies specific for and/or nucleic acid are
```

```
CC   used for the prevention and/or treatment of breast cancer. Antibodies
CC   against may also be used for screening for and/or diagnosis of breast
CC   cancer in a patient. The method for monitoring/assessing breast cancer
CC   treatment in a patient and for the identification of metastatic breast
CC   cancer cells in samples from a patient. This sequence encodes breast
CC   cancer cell associated protein 7 (BCMP 7), located on chromosome
CC   7p21.3, described in the method of the invention.
XX
SQ   Sequence 543 BP; 170 A; 135 C; 107 G; 131 T; 0 other;


alignment_scores:
            Quality:  912.00              Length:    178
              Ratio:  5.124                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x AAS13480   ..

 Align seg 1/1  to: AAS13480  from: 1  to: 543

      29 ArgValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAl 45
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       2 AGAGTTGCCATGGAGAAAATCCCAGTGTCAGCATTCTTGCTCCTTGTGGC 51

      45 aLeuSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysL 62
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      52 CCTCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAA 101

      62 ysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGly 78
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     102 AGGACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGT 151

      79 TrpGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLy 95
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     152 TGGGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAA 201

      95 sSerLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluC 112
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     202 ATCCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGT 251

     112 ysProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIle 128
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     252 GCCCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATC 301

     129 GlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrTh 145
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     302 CAGAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAAC 351

     145 rAspLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheV 162
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     352 TGACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTG 401

     162 alAspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsn 178
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     402 TTGACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAAC 451

     179 ArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMe 195
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     452 CGTCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACAT 501

     195 tLysLysAlaLeuLysLeuLeuLysThrGluLeu 206
         |||||||||||||||||||||||||||||||||
     502 GAAGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 535

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:AAV29048

seq_documentation_block:
ID   AAV29048 standard; cDNA; 525 BP.
XX
AC   AAV29048;
XX
DT   21-AUG-1998  (first entry)
XX
DE   Open reading frame human protein comprising secretory signal 9.
XX
KW   Human protein; secretory signal; nutritional source; cytokine;
KW   immunity;haematopoiesis; activin; inhibin; tumour; chemotactic;
KW   chemokinetic; thrombolytic; anti-inflammatory; inhibition;
KW   stomach cancer cell; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..525
FT                   /*tag=  a
FT                   /product= "human protein comprising secretory
FT                            signal"
XX
PN   WO9811217-A2.
XX
PD   19-MAR-1998.
XX
PF   12-SEP-1997;   97WO-JP03239.
XX
PR   13-SEP-1996;   96JP-0243060.
XX
PA   (PROT-) PROTEGENE INC.
PA   (SAGA ) SAGAMI CHEM RES CENTRE.
XX
PI   Kato S,  Kobayashi M,  Sekine S,  Yamaguchi T;
XX
DR   WPI; 1998-207380/18.
DR   P-PSDB; AAW37872.
XX
PT   Human proteins with secretory signal sequences - used to treat
PT   immune deficiencies, infections, tumours, and haematopoietic
PT   disorders, etc.
XX
PS   Claim 3; Pages 88; 131pp; English.
XX
CC   This is the nucleotide sequence of the open reading frame of a novel
CC   human protein comprising a secretory signal (AAV29047), isolated from
CC   stomach cancer cells.  Its proteins can be used as nutritional sources
CC   or supplements.  The proteins may also have cytokine functions,
CC   immune modulating  functions, haematopoiesis regulating activity,
CC   activin/inhibin regulating activity, chemotactic/chemokinetic activity,
CC   haemostatic and thrombolytic activity, receptor/ligand activity,
CC   anti-inflammatory activity, tumour inhibition activity.
XX
SQ   Sequence 525 BP; 162 A; 131 C; 103 G; 129 T; 0 other;


alignment_scores:
            Quality:  899.00              Length:    175
              Ratio:  5.137                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x AAV29048   ..

 Align seg 1/1  to: AAV29048  from: 1  to: 525

      32 MetGluLysIleProValSerAlaPheLeuLeuLeuValAlaLeuSerTy 48
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCTCCTA 50

      48 rThrLeuAlaArgAspThrThrValLysProGlyAlaLysLysAspThrL 65
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 CACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGACACAA 100

      65 ysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTrpGlyAsp 81
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 AGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGGGTGAC 150

      82 GlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerLysTh 98
```

```
     ||||||||||||||||||||||||||||||||||||||||||||||||||
151 CAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAATCCAAGAC 200

 98 rSerAsnLysProLeuMetIleIleHisHisLeuAspGluCysProHisS 115
    ||||||||||||||||||||||||||||||||||||||||||||||||||
201 AAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGCCCACACA 250

115 erGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGlnLysLeu 131
    ||||||||||||||||||||||||||||||||||||||||||||||||||
251 GTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCAGAAATTG 300

132 AlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrAspLysHi 148
    ||||||||||||||||||||||||||||||||||||||||||||||||||
301 GCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTGACAAACA 350

148 sLeuSerProAspGlyGlnTyrValProArgIleMetPheValAspProS 165
    ||||||||||||||||||||||||||||||||||||||||||||||||||
351 CCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTTGACCCAT 400

165 erLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnArgLeuTyr 181
    ||||||||||||||||||||||||||||||||||||||||||||||||||
401 CTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAACCGTCTCTAT 450

182 AlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLysAl 198
    ||||||||||||||||||||||||||||||||||||||||||||||||||
451 GCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACATGAAGAAAGC 500

198 aLeuLysLeuLeuLysThrGluLeu 206
    |||||||||||||||||||||||||
501 TCTCAAGTTGCTGAAGACTGAATTG 525
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAZ24578

seq_documentation_block:
ID AAZ24578 standard; cDNA; 793 BP.
XX
AC AAZ24578;
XX
DT 07-DEC-1999 (first entry)
XX
DE Human lung tumor associated polynucleotide.
XX
KW Human; lung tumor; lung cancer; T cell stimulation; ss.
XX
OS Homo sapiens.
XX
PN WO9947674-A2.
XX
PD 23-SEP-1999.
XX
PF 17-MAR-1999; 99WO-US05798.
XX
PR 18-MAR-1998; 98US-0040802.
PR 18-MAR-1998; 98US-0040984.
PR 27-JUL-1998; 98US-0123912.
PR 27-JUL-1998; 98US-0123933.
XX
PA (CORI-) CORIXA CORP.
XX
PI Reed SG, Wang T;
XX
DR WPI; 1999-571839/48.
XX
PT New isolated lung tumor polynucleotides, used to develop products for
PT the treatment, prevention and monitoring the progression of lung cancer
PT -
XX
PS Claim 1; Page 90; 148pp; English.
XX
CC The invention provides isolated human lung tumor nucleic acids and
CC polypeptides. The polypeptides can be used for the treatment of lung
CC cancer. The polypeptides and polynucleotides can be used to stimulate T
CC cells or antigen presenting cells for use in the treatment of lung
CC cancer. The polypeptides and monoclonal antibodies specific for the
CC polypeptides can also be used to inhibit the development of lung cancer.
CC Agents which bind the polypeptides can be used for detecting lung cancer
CC and for monitoring the progression of lung cancer.
XX
SQ Sequence 793 BP; 246 A; 172 C; 148 G; 211 T; 16 other;

alignment_scores:
Quality: 856.50 Length: 190
Ratio: 4.706 Gaps: 1
Percent Similarity: 95.789 Percent Identity: 93.684

alignment_block:
US-09-674-266A-181 x AAZ24578 ..

Align seq 1/1 to: AAZ24578 from: 1 to: 793

```
 14 HisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerArgVa 30
    ||||||||||||||||||||||||||||||||||||||||||||||||||
  2 CATCCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAGAGT 51

 30 lAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAlaLeuS 47
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 52 TGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCT 101

 47 erTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLysAsp 63
    ||||||||||||||||||||||||||||||||||||||||||||||||||
102 CCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGAC 151

 64 ThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTrpGl 80
    ||||||||||||||||||||||||||||||||||||||||||||||||||
152 ACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGGG 201

 80 yAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerL 97
    ||||||||||||||||||||||||||||||||||||||||||||||||||
202 TGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAATCCA 251

 97 ysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCysPro 113
    ||||||||||||||||||||||||||||||||||||||||||||||||||
252 AGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGCCCA 301

114 HisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGlnLy 130
    ||||||  ||||||||||||||||||||||||||||||||||||||||||
302 CACAGTCNAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCAGAA 351

130 sLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrAspL 147
    ||||||||||||||||||||||||||||||||||||||||||||||||||
352 ATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTGACA 401

147 ysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheValAsp 163
    ||||||||||||||||||||||||  ||||||||||||||||||||||||
402 AACACCTTTCTCCTGATGGCCAGTATGT.CCCAGGATTATGTTTGTTGAC 450

164 ProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnArgLe 180
    ||||||||||||||:::|||||||||  ||||||||||||||||||||||
451 CCATCTCTGACAGTTGAAGCCGATATCCTGGGAAGATATTCNAACCGTCT 500

180 uTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetLysL 197
    ||||||||||:::  |||||||||        |||  :::  :::|||||
501 CTATGCTTACAAACT.GCAGATACGCTCTGTTGCTTGACACAT...GAAA 546

197 ysAlaLeuLysLeuLeuLys 203
    ||||||||||||||||||||
547 AAGCTCTCAAGTTGCTNAAA 566
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAC65817

seq_documentation_block:
ID AAC65817 standard; cDNA; 793 BP.
XX
AC AAC65817;
XX

```
DT   21-FEB-2001  (first entry)
XX
DE   Human lung cancer-associated cDNA LST-S1-A-10A.
XX
KW   Lung cancer; therapy; treatment; human; tumor; immunogenic; cytostatic;
KW   vaccine; detection; ss.
XX
OS   Homo sapiens.
XX
PN   WO200061612-A2.
XX
PD   19-OCT-2000.
XX
PF   03-APR-2000; 2000WO-US08896.
XX
PR   02-APR-1999;   99US-0285479.
PR   17-DEC-1999;   99US-0466396.
PR   30-DEC-1999;   99US-0476496.
PR   10-JAN-2000; 2000US-0480884.
PR   22-FEB-2000; 2000US-0510376.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Wang T,  Fan L;
XX
DR   WPI; 2000-628399/60.
XX
PT   Isolated polypeptide comprising an immunogenic portion of a lung tumor
PT   protein is used for detecting and monitoring progression of lung cancer
PT   in a patient -
XX
PS   Claim 1a; Page 123; 261pp; English.
XX
CC   This invention describes a novel isolated polypeptide (I) which
CC   comprising an immunogenic portion of a lung tumor protein or variant (P2)
CC   which have cytostatic activity. The polypeptides and polynucleotides are
CC   used in compositions and vaccines to inhibit the development of cancer,
CC   especially lung cancer, in a patient. Methods described in the invention
CC   can be used to monitor the progression of a cancer by carrying out the
CC   detection at subsequent time points and comparing the results from the
CC   different time points. CD4+ and/or CD8+ T-Cells isolated from a patient
CC   are treated with P2, polynucleotides encoding P2 or antigen presenting
CC   cells expressing P2 and then administered to the patient to inhibit
CC   development of cancer.
XX
SQ   Sequence 793 BP; 246 A; 172 C; 148 G; 211 T; 16 other;


alignment_scores:
            Quality:  856.50              Length:    190
              Ratio:  4.706                Gaps:      1
 Percent Similarity:  95.789   Percent Identity: 93.684

alignment_block:
 US-09-674-266A-181 x AAC65817   ..

 Align seg 1/1  to: AAC65817  from: 1  to: 793

      14 HisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerArgVa 30
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       2 CATCCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAGAGT 51

      30 lAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAlaLeuS 47
         |||||||||||||||||||||||||||||||||||||||||||||||||
      52 TGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCT 101

      47 erTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLysAsp 63
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     102 CCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGAC 151

      64 ThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTrpGl 80
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     152 ACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGGG 201

      80 yAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerL 97
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     202 TGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAATCCA 251

      97 ysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCysPro 113
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     252 AGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGCCCA 301

     114 HisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGlnLy 130
         ||||||   ||||||||||||||||||||||||||||||||||||||||||
     302 CACAGTCNAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCAGAA 351

     130 sLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrAspL 147
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     352 ATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTGACA 401

     147 ysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheValAsp 163
         |||||||||||||||||||||||||   |||||||||||||||||||||
     402 AACACCTTTCTCCTGATGGCCAGTATGT.CCCAGGATTATGTTTGTTGAC 450

     164 ProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnArgLe 180
         ||||||||||||||:::|||||||||   ||||||||||||||||||||
     451 CCATCTCTGACAGTTGAAGCCGATATCCTGGGAAGATATTCNAACCGTCT 500

     180 uTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetLysL 197
         ||||||||||:::   |||||||||        |||   :::   :::|
     501 CTATGCTTACAAACT.GCAGATACGCTCTGTTGCTTGACACAT...GAAA 546

     197 ysAlaLeuLysLeuLeuLys 203
         ||||||||||||||||||
     547 AAGCTCTCAAGTTGCTNAAA 566
```

OM of: US-09-674-266A-181 to: Issued_Patents_NA:* out_format : pfs

Date: Oct 8, 2002 10:03 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09674266/runat_08102002_093325_26183/app_query.fasta_1.271
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=p2n.rni
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
-LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000
-XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
-MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
-THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09674266_@CGN1_1_46 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-674-266A-181
Query length: 206
Database: Issued_Patents_NA:*
Database sequences: 383533
Database length: 122816752
Search time (sec): 46.580000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-1	+	991.00	2195.32	1.4e-114	875	!
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-247-155-61	+	961.00	2120.07	2.1e-110	1689	!
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-247-155-149	+	914.00	2021.63	6.4e-105	940	!
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-040-984-78	+	856.50	1894.61	7.7e-98	793	!
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-123-912-78	+	856.50	1894.61	7.7e-98	793	!
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-221-298-32	+	683.00	1513.40	1.3e-76	401	!
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-35	+	642.00	1421.83	1.7e-71	386	!
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-5	+	555.00	1218.86	3.3e-60	732	!
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-36	+	487.50	1075.46	3.2e-52	373	!
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-9	+	458.50	1007.11	2.1e-48	489	!
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-10	+	376.50	822.65	3.9e-38	506	!
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-37	+	253.00	559.46	1.8e-23	158	!
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-030-607-209	+	252.00	557.14	2.4e-23	159	!
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-439-313-209	+	252.00	557.14	2.4e-23	159	!
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-3	+	222.00	463.43	4.0e-18	1423	!
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-38	+	183.50	405.27	6.9e-15	136	!
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-19	+	158.00	338.43	3.7e-11	302	!
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-41	+	147.50	314.55	7.8e-10	310	!
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-40	-	129.00	272.99	1.6e-07	311	!
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-11	+	104.00	216.20	0.0002	329	!
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-12	+	104.00	216.20	0.0002	329	!
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-13	+	104.00	216.20	0.0002	329	!
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-14	+	104.00	216.20	0.0002	329	!
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-15	+	101.00	209.43	0.0006	330	!
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-22	+	95.50	201.69	0.0015	225	!
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-534-638-1	+	88.50	140.52	3.87	9840	!
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-18	+	87.50	174.01	0.0528	505	!
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-16	+	82.00	166.93	0.1308	326	!
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-745-206A-14	-	79.50	136.95	6.12	2470	!
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-311-363-14	-	79.50	136.95	6.12	2470	!
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-745-206A-12	-	79.50	127.39	20.85	5467	!
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-311-363-12	-	79.50	127.39	20.85	5467	!
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-455-543A-8	-	79.50	124.12	31.71	7175	!
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-193-078B-8	-	79.50	124.12	31.71	7175	!
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-223-305C-8	-	79.50	124.12	31.71	7175	!
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-149-097D-8	-	79.50	124.12	31.71	7175	!
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-949-386-8	-	79.50	124.12	31.71	7175	!
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-450-562-8	-	79.50	124.12	31.71	7175	!
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-984-709A-8	-	79.50	124.12	31.71	7175	!
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-268-163-7	-	79.50	124.12	31.73	7177	!
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-713-118-1	-	79.50	123.97	32.34	7266	
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-09-452-007-1	-	79.50	123.97	32.34	7266	
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-455-543A-7	-	79.50	123.81	33.00	7362	
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-193-078B-7	-	79.50	123.81	33.00	7362	
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-223-305C-7	-	79.50	123.81	33.00	7362	

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-1

seq_documentation_block:
; Sequence 1, Application US/08916576B
; Patent No. 6171816
; GENERAL INFORMATION:
; APPLICANT: YU, GUO-LIANG
; APPLICANT: DILLON, PATRICK J.
; APPLICANT: EBNER, REINHARD
; APPLICANT: ENDRESS, GREGORY A.
; TITLE OF INVENTION: NOVEL HUMAN GROWTH FACTORS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: US
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/916,576B
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,347
; FILING DATE: 23-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0500001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 875 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 71..595
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 131..595
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 71..130
US-08-916-576B-1

alignment_scores:
Quality: 991.00 Length: 194
Ratio: 5.135 Gaps: 0
Percent Similarity: 99.485 Percent Identity: 99.485

alignment_block:
US-09-674-266A-181 x US-08-916-576B-1 ..

Align seg 1/1 to: US-08-916-576B-1 from: 1 to: 875

```
     13 ProHisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerAr 29
        |||   ||||||||||||||||||||||||||||||||||||||||||||
     14 CCGATTCCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAG 63

     29 gValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAlaL 46
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     64 AGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCC 113

     46 euSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLys 62
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    114 TCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAG 163

     63 AspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTr 79
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    164 GACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTG 213

     79 pGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysS 96
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    214 GGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAAT 263

     96 erLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCys 112
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    264 CCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGC 313

    113 ProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGl 129
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    314 CCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCA 363

    129 nLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrA 146
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    364 GAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTG 413

    146 spLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheVal 162
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    414 ACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTT 463

    163 AspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnAr 179
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    464 GACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAATCG 513

    179 gLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetL 196
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    514 TCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACATGA 563

    196 ysLysAlaLeuLysLeuLeuLysThrGluLeu 206
        |||||||||||||||||||||||||||||||
    564 AGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 595
```

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-247-155-61

seq_documentation_block:

```
; Sequence 61, Application US/09247155A
; Patent No. 6312922
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, Jean-Baptiste
;  APPLICANT: Duclert, Aymeric
;  APPLICANT: Bougueleret, Lydie
;  TITLE OF INVENTION: Complementary DNAs
;  FILE REFERENCE: GENSET.021A
;  CURRENT APPLICATION NUMBER: US/09/247,155A
;  CURRENT FILING DATE: 1999-02-09
;  EARLIER APPLICATION NUMBER: 60/074,121
;  EARLIER FILING DATE: 1998-02-09
;  EARLIER APPLICATION NUMBER: 60/081,563
;  EARLIER FILING DATE: 1998-04-13
;  EARLIER APPLICATION NUMBER: 60/096,116
;  EARLIER FILING DATE: 1998-08-10
;  EARLIER APPLICATION NUMBER: 60/099,273
;  EARLIER FILING DATE: 1998-10-04
;  NUMBER OF SEQ ID NOS: 182
;  SOFTWARE: Patent.pm
; SEQ ID NO 61
;   LENGTH: 1689
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 51..575
;   FEATURE:
;   NAME/KEY: sig_peptide
;   LOCATION: 51..110
;   OTHER INFORMATION: Von Heijne matrix
;   OTHER INFORMATION: score 11.2
;   OTHER INFORMATION: seq AFLLLVALSYTLA/RD
;   FEATURE:
;   NAME/KEY: polyA_signal
;   LOCATION: 1653..1658
;   FEATURE:
;   NAME/KEY: polyA_site
;   LOCATION: 1674..1689
US-09-247-155-61
```

alignment_scores:

```
            Quality:  961.00              Length:    195
              Ratio:  5.085                 Gaps:      1
 Percent Similarity:  96.923   Percent Identity: 96.923
```

alignment_block:

US-09-674-266A-181 x US-09-247-155-61 ..

Align seg 1/1 to: US-09-247-155-61 from: 1 to: 1689

```
     12 GlyProHisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSe 28
        ||||||||||||||||||||||||||||||||||
      9 GGACCGCATCCTAGCCGCCGACTCACACAAGGC..................  41

     28 rArgValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValA 45
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     42 .AGAGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGG 90

     45 laLeuSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLys 61
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     91 CCCTCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAA 140

     62 LysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGl 78
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    141 AAGGACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGG 190

     78 yTrpGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrL 95
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    191 TTGGGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATA 240

     95 ysSerLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGlu 111
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    241 AATCCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAG 290

    112 CysProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIl 128
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    291 TGCCCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAAT 340

    128 eGlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrT 145
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    341 CCAGAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAA 390

    145 hrAspLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPhe 161
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    391 CTGACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTT 440

    162 ValAspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAs 178
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    441 GTTGACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAA 490

    178 nArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnM 195
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    491 TCGTCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACA 540
```

```
    195 etLysLysAlaLeuLysLeuLeuLysThrGluLeu 206
        ||||||||||||||||||||||||||||||||||||
    541 TGAAGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 575

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-247-155-149

seq_documentation_block:
; Sequence 149, Application US/09247155A
; Patent No. 6312922
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, Jean-Baptiste
;  APPLICANT: Duclert, Aymeric
;  APPLICANT: Bougueleret, Lydie
;  TITLE OF INVENTION: Complementary DNAs
;  FILE REFERENCE: GENSET.021A
;  CURRENT APPLICATION NUMBER: US/09/247,155A
;  CURRENT FILING DATE: 1999-02-09
;  EARLIER APPLICATION NUMBER: 60/074,121
;  EARLIER FILING DATE: 1998-02-09
;  EARLIER APPLICATION NUMBER: 60/081,563
;  EARLIER FILING DATE: 1998-04-13
;  EARLIER APPLICATION NUMBER: 60/096,116
;  EARLIER FILING DATE: 1998-08-10
;  EARLIER APPLICATION NUMBER: 60/099,273
;  EARLIER FILING DATE: 1998-10-04
;  NUMBER OF SEQ ID NOS: 182
;  SOFTWARE: Patent.pm
; SEQ ID NO 149
;   LENGTH: 940
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 177..569
;   FEATURE:
;   NAME/KEY: sig_peptide
;   LOCATION: 177..236
;   OTHER INFORMATION: Von Heijne matrix
;   OTHER INFORMATION: score 11.1999998092651
;   OTHER INFORMATION: seq AFLLLVALSYTLA/RD
;   FEATURE:
;   NAME/KEY: polyA_site
;   LOCATION: 931..939
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 482
;   OTHER INFORMATION: n=a, g, c or t
US-09-247-155-149


alignment_scores:
            Quality:  914.00              Length:    207
              Ratio:  4.811                 Gaps:      2
 Percent Similarity:  91.787   Percent Identity: 90.821

alignment_block:
 US-09-674-266A-181 x US-09-247-155-149   ..

 Align seg 1/1  to: US-09-247-155-149  from: 1  to: 940

       1 ArgLeuSerCysAlaGlyThrLeuSerGlySerGlyProHisProSerAr 17
         ||||||||||||||||||||||||   :::   ||||||||||||||||||
     103 AGACTCAGCTGTGCTGGCACACTCAGAAGCTT.GGACCGCATCCTAGCCG 151

      17 gArgLeuThrGlnGlyArgTrpValArgLysSerArgValAlaMetGluL 34
         ||||||||||||||||                 ||||||||||||||||
     152 CCGACTCACACAAGGC.................AGAGTTGCCATGGAAA 183

      34 ysIleProValSerAlaPheLeuLeuLeuValAlaLeuSerTyrThrLeu 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     184 AAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCTCCTACACTCTG 233

      51 AlaArgAspThrThrValLysProGlyAlaLysLysAspThrLysAspSe 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     234 GCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGACACAAAGGACTC 283

      67 rArgProLysLeuProGlnThrLeuSerArgGlyTrpGlyAspGlnLeuI 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     284 TCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGGGTGACCAACTCA 333

      84 leTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerLysThrSerAsn 100
         ||||||||||||||||||   :::|||   ||||||||||||||||||
     334 TCTGGACACARACATATGAARAARCTCTATWTAAATCCAARACAAGCAAC 383

     101 LysProLeuMetIleIleHisHisLeuAspGluCysProHisSerGlnAl 117
         ||||||||||||||||||||||||||||||   ||||||||||||||||
     384 AAACCCTTGATGATTATTCATCACTTGGATGADTGCCCACACAGTCAAGC 433

     117 aLeuLysLysValPheAlaGluAsnLysGluIleGlnLysLeuAlaGluG 134
         ||||||||||   ||||||||||||||   ||||||||||||||   |
     434 TTTAAAAAAAAKTGTTTGCTGAAAATAAARAAATCCAGAAATTGGCARANC 483

     134 lnPheValLeuLeuAsnLeuValTyrGluThrThrAspLysHisLeuSer 150
         ||||||||   ||||||||||||||||||||||||||||||||||||||
     484 AGTTTGTCCYCCTCAATCTGGTTTATGAAACAACTGACAAACACCTTTCT 533

     151 ProAspGlyGlnTyrVal.ProArgIleMetPheValAspProSerLeuT 167
         ||||||||||||||||    ||||||||||||||||||||||||||||||
     534 CCTGATGGCCAATATKTCCCCCMGGATTATGTTTGTTGACCCATCTCTGA 583

     167 hrValArgAlaAspIleThrGlyArgTyrSerAsnArgLeuTyrAlaTyr 183
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     584 CAGTTAGAGCCGATATCACTGGAAGATATTCAAAYCGTCTCTATGCTTAC 633

     184 GluProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLysAlaLeuLy 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     634 GAACCTGCAGATACAGCTCTGTTGCTTGACAACATGAAGAAAGCTCTCAA 683

     200 sLeuLeuLysThrGluLeu 206
         |||||||||||||||||||
     684 GTTGCTGAAGACTGAATTG 702

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-040-984-78

seq_documentation_block:
; Sequence 78, Application US/09040984
; Patent No. 6210883
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    APPLICANT:  Wang, TongTong
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR DIAGNOSIS
;    TITLE OF INVENTION:  OF LUNG CANCER
;    NUMBER OF SEQUENCES:  86
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/040,984
;      FILING DATE:  18-MAR-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.456
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-622-4900
```

```
;      TELEFAX:  206-282-6031
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  78:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  793 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-040-984-78


alignment_scores:
            Quality:  856.50              Length:    190
              Ratio:  4.706                 Gaps:      1
 Percent Similarity:  95.789   Percent Identity: 93.684

alignment_block:
 US-09-674-266A-181 x US-09-040-984-78   ..

 Align seg 1/1  to: US-09-040-984-78  from: 1  to: 793

      14 HisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerArgVa 30
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       2 CATCCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAGAGT 51

      30 lAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAlaLeuS 47
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      52 TGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCT 101

      47 erTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLysAsp 63
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     102 CCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGAC 151

      64 ThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTrpGl 80
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     152 ACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGGG 201

      80 yAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerL 97
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     202 TGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAATCCA 251

      97 ysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCysPro 113
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     252 AGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGCCCA 301

     114 HisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGlnLy 130
         ||||||   ||||||||||||||||||||||||||||||||||||||||||
     302 CACAGTCNAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCAGAA 351

     130 sLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrAspL 147
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     352 ATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTGACA 401

     147 ysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheValAsp 163
         ||||||||||||||||||||||||   ||||||||||||||||||||
     402 AACACCTTTCTCCTGATGGCCAGTATGT.CCCAGGATTATGTTTGTTGAC 450

     164 ProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnArgLe 180
         ||||||||||||||:::|||||||||   ||||||||||||||||||||
     451 CCATCTCTGACAGTTGAAGCCGATATCCTGGGAAGATATTCNAACCGTCT 500

     180 uTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetLysL 197
         ||||||||||:::   |||||||||         |||   :::   :::|
     501 CTATGCTTACAAACT.GCAGATACGCTCTGTTGCTTGACACAT...GAAA 546

     197 ysAlaLeuLysLeuLeuLys 203
         ||||||||||||||||||||
     547 AAGCTCTCAAGTTGCTNAAA 566

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-123-912-78

seq_documentation_block:
; Sequence 78, Application US/09123912A

; Patent No. 6312695
; GENERAL INFORMATION:
;  APPLICANT:  Reed, Steven G.
;  APPLICANT:  Wang, Tongtong
;  TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR THERAPY OF LUNG CANCER
;  FILE REFERENCE:  210121.455C1
;  CURRENT APPLICATION NUMBER: US/09/123,912A
;  CURRENT FILING DATE:  1998-07-27
;  PRIOR APPLICATION NUMBER:  09/040,802
;  PRIOR FILING DATE:  1998-03-18
;  NUMBER OF SEQ ID NOS:  114
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 78
;   LENGTH: 793
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (309)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (492)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (563)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (657)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (660)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (703)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (708)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (710)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (711)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (732)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (740)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (748)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (758)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (762)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (765)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (787)
;   OTHER INFORMATION: Where n is a, c, g or t
US-09-123-912-78


alignment_scores:
            Quality:  856.50              Length:    190
              Ratio:  4.706                 Gaps:      1
 Percent Similarity:  95.789   Percent Identity: 93.684
```

```
alignment_block:
 US-09-674-266A-181 x US-09-123-912-78   ..

 Align seg 1/1  to: US-09-123-912-78  from: 1  to: 793

      14 HisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerArgVa 30
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
       2 CATCCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAGAGT 51

      30 lAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAlaLeuS 47
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      52 TGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCT 101

      47 erTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLysAsp 63
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     102 CCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGAC 151

      64 ThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTrpGl 80
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     152 ACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGGG 201

      80 yAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerL 97
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     202 TGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAATCCA 251

      97 ysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCysPro 113
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     252 AGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGCCCA 301

     114 HisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGlnLy 130
         ||||||   ||||||||||||||||||||||||||||||||||||||||||
     302 CACAGTCNAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCAGAA 351

     130 sLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrAspL 147
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     352 ATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTGACA 401

     147 ysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheValAsp 163
         ||||||||||||||||||||||||||   ||||||||||||||||||||
     402 AACACCTTTCTCCTGATGGCCAGTATGT.CCCAGGATTATGTTTGTTGAC 450

     164 ProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnArgLe 180
         ||||||||||||||:::|||||||||   ||||||||||||||||||||
     451 CCATCTCTGACAGTTGAAGCCGATATCCTGGGAAGATATTCNAACCGTCT 500

     180 uTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetLysL 197
         ||||||||||:::   |||||||||         |||   :::   :::|
     501 CTATGCTTACAAACT.GCAGATACGCTCTGTTGCTTGACACAT...GAAA 546

     197 ysAlaLeuLysLeuLeuLys 203
         ||||||||||||||||||||
     547 AAGCTCTCAAGTTGCTNAAA 566

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-221-298-32

seq_documentation_block:
; Sequence 32, Application US/09221298
; Patent No. 6284241
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY AND DIAGNOSIS
;  TITLE OF INVENTION: OF COLON CANCER
;  FILE REFERENCE: 210121.471
;  CURRENT APPLICATION NUMBER: US/09/221,298
;  CURRENT FILING DATE: 1998-12-23
;  NUMBER OF SEQ ID NOS: 112
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
;   LENGTH: 401
;   TYPE: DNA
;   ORGANISM: Human
US-09-221-298-32
```

```
alignment_scores:
           Quality:  683.00              Length:    134
             Ratio:  5.174                 Gaps:      1
 Percent Similarity:  98.507   Percent Identity: 98.507

alignment_block:
 US-09-674-266A-181 x US-09-221-298-32   ..

 Align seg 1/1  to: US-09-221-298-32  from: 1  to: 401

      57 LysProGlyAlaLysLysAspThrLysAspSerArgProLysLeuProGl 73
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
       2 AAACCTGGAGCCAAAAAGGACACAAAGGACTCTCGACCCAAACTGCCCCA 51

      73 nThrLeuSerArgGlyTrpGlyAspGlnLeuIleTrpThrGlnThrTyrG 90
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      52 GACCCTCTCCAGAGGTTGGGGTGACCAACTCATCTGGACTCAGACATATG 101

      90 luGluAlaLeuTyrLysSerLysThrSerAsnLysProLeuMetIleIle 106
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     102 AAGAAGCTCTATATAAATCCAAGACAAGCAACAAACCCTTGATGATTATT 151

     107 HisHisLeuAspGluCysProHisSerGlnAlaLeuLysLysValPheAl 123
         |||||||||   ||||||||||||||||||||||||||||||||||||||||
     152 CATCACTTGGGTGAGTGCCCACACAGTCAAGCTTTAAAGAAAGTGTTTGC 201

     123 aGluAsnLysGluIleGlnLysLeuAlaGluGlnPheValLeuLeuAsnL 140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     202 TGAAAATAAAGAAATCCAGAAATTGGCAGAGCAGTTTGTCCTCCTCAATC 251

     140 euValTyrGluThrThrAspLysHisLeuSerProAspGlyGlnTyrVal 156
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     252 TGGTTTATGAAACAACTGACAAACACCTTTCTCCTGATGGCCAGTATGTC 301

     157 ProArgIleMetPheValAspProSerLeuThrValArgAla.AspIleT 173
         |||||||||||||||||||||||||||||||||||||| |||||||
     302 CCCAGGATTATGTTTGTTGACCCATCTCTGACAGTTAGAGCCCGATATCA 351

     173 hrGlyArgTyrSerAsnArgLeuTyrAlaTyrGluProAlaAspThrAla 189
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     352 CTGGAAGATATTCAAACCGTCTCTATGCTTACGAACCTGCAGATACAGCT 401

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-35

seq_documentation_block:
; Sequence 35, Application US/08916576B
; Patent No. 6171816
;  GENERAL INFORMATION:
;    APPLICANT:  YU, GUO-LIANG
;    APPLICANT:  DILLON, PATRICK J.
;    APPLICANT:  EBNER, REINHARD
;    APPLICANT:  ENDRESS, GREGORY A.
;    TITLE OF INVENTION:  NOVEL HUMAN GROWTH FACTORS
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  US
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/916,576B
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/024,347
```

```
;       FILING DATE:  23-AUG-1996
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  STEFFE, ERIC K.
;       REGISTRATION NUMBER:  36,688
;       REFERENCE/DOCKET NUMBER:  1488.0500001
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202) 371-2600
;       TELEFAX:  (202) 371-2540
;   INFORMATION FOR SEQ ID NO:  35:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  386 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-916-576B-35


alignment_scores:
            Quality:  642.00              Length:    128
              Ratio:  5.016                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 97.656

alignment_block:
 US-09-674-266A-181 x US-08-916-576B-35   ..

 Align seg 1/1  to: US-08-916-576B-35  from: 1  to: 386

      61 LysLysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerAr 77
         :::::|||||||||||||||||||||||||||||||||||||||||||||
       2 CGACAGGACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAG 51

      77 gGlyTrpGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuT 94
         ||||||||||||   ||||||||||||||||||||||||||||||||||||
      52 AGGTTGGGGTGACAA.CTCATCTGGACTCAGACATATGAAGAAGCTCTAT 100

      94 yrLysSerLysThrSerAsnLysProLeuMetIleIleHisHisLeuAsp 110
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 ATAAATCCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGAT 150

     111 GluCysProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGl 127
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 GAGTGCCCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGA 200

     127 uIleGlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluT 144
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 AATCCAGAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAA 250

     144 hrThrAspLysHisLeuSerProAspGlyGlnTyrValProArgIleMet 160
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 CAACTGACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATG 300

     161 PheValAspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSe 177
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     301 TTTGTTGACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTC 350

     177 rAsnArgLeuTyrAlaTyrGluProAlaAspThr 188
         |||||||||||||||||||||||||||||||||
     351 AAATCGTCTCTATGCTTACGAACCTGCAGATACA 384

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-5

seq_documentation_block:
; Sequence 5, Application US/08916576B
; Patent No. 6171816
;  GENERAL INFORMATION:
;     APPLICANT:  YU, GUO-LIANG
;     APPLICANT:  DILLON, PATRICK J.
;     APPLICANT:  EBNER, REINHARD
;     APPLICANT:  ENDRESS, GREGORY A.
;     TITLE OF INVENTION:  NOVEL HUMAN GROWTH FACTORS
;     NUMBER OF SEQUENCES:  45
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
;       STREET:  1100 NEW YORK AVENUE, SUITE 600
;       CITY:  WASHINGTON
;       STATE:  DC
;       COUNTRY:  US
;       ZIP:  20005-3934
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/916,576B
;       FILING DATE:
;       CLASSIFICATION:  536
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 60/024,347
;       FILING DATE:  23-AUG-1996
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  STEFFE, ERIC K.
;       REGISTRATION NUMBER:  36,688
;       REFERENCE/DOCKET NUMBER:  1488.0500001
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202) 371-2600
;       TELEFAX:  (202) 371-2540
;   INFORMATION FOR SEQ ID NO:  5:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  732 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  49..546
;     FEATURE:
;       NAME/KEY:  mat_peptide
;       LOCATION:  118..546
;     FEATURE:
;       NAME/KEY:  sig_peptide
;       LOCATION:  49..117
US-08-916-576B-5


alignment_scores:
            Quality:  555.00              Length:    167
              Ratio:  3.801                 Gaps:      4
 Percent Similarity:  87.425   Percent Identity: 64.671

alignment_block:
 US-09-674-266A-181 x US-08-916-576B-5   ..

 Align seg 1/1  to: US-08-916-576B-5  from: 1  to: 732

      41 LeuLeuLeuValAlaLeuSerTyrThrLeuAlaArgAspThrThrValLy 57
         ||||||||||||:::::|||   :::||||||         :::::||
      79 CTCTTACTCGTCACAGTTTCTTCCAACCTTGCC......ATTGCAATAAA 122

      57 sProGlyAlaLysLysAspThrLysAspSerArgProLysLeuProGlnT 74
         |                     |||::::::||||||      |||||||
     123 A.....................AAGGAAAAGAGGCCT......CCTCAGA 145

      74 hrLeuSerArgGlyTrpGlyAspGlnLeuIleTrpThrGlnThrTyrGlu 90
         ||||||||||||||||||||||::::::   |||:::|||||||||||
     146 CACTCTCAAGAGGATGGGGAGATGACATCACTTGGGTACAAACTTATGAA 195

      91 GluAlaLeuTyrLysSerLysThrSerAsnLysProLeuMetIleIleHi 107
         |||:::|||:::   ::::::   |||:::|||||||||||:::||||||
     196 GAAGGTCTCTTTTATGCTCAAAAAAGTAAGAAGCCATTAATGGTTATTCA 245

     107 sHisLeuAspGluCysProHisSerGlnAlaLeuLysLysValPheAlaG 124
         |||||||:::::::|||   :::||||||||||||||||||||||||||:
     246 TCACCTGGAGGATTGTCAATACTCTCAAGCACTAAAGAAAGTATTTGCCC 295
```

```
     124 luAsnLysGluIleGlnLysLeuAlaGlu...GlnPheValLeuLeuAsn 139
         ::|||:::|||||||||:::::::|||:::   :::|||::::::|||||
     296 AAAATGAAGAAATACAAGAAATGGCTCAGAATAAGTTCATCATGCTAAAC 345

     140 LeuValTyrGluThrThrAspLysHisLeuSerProAspGlyGlnTyrVa 156
         |||::::::|||||||||||||||:::|||||||||||||||||||||||
     346 CTTATGCATGAAACCACTGATAAGAATTTATCACCTGATGGGCAATATGT 395

     156 lProArgIleMetPheValAspProSerLeuThrValArgAlaAspIleT 173
         ||||||||||||||||||||||||||||||||||||||||||||||||||:
     396 GCCTAGAATCATGTTTGTAGACCCTTCTTTAACAGTTAGAGCTGACATAG 445

     173 hrGlyArgTyrSerAsnArgLeuTyrAlaTyrGluProAlaAspThrAla 189
         ::|||||||||||||||||||||||:::|||||||||   |||
     446 CTGGAAGATACTCTAACAGATTGTACACATATGAGCCTCGGGATTTACCC 495

     190 LeuLeuLeuAspAsnMetLysLysAlaLeuLysLeuLeuLysThrGluLe 206
         ||||||::::::|||||||||||||||||||:::|||::::::::::|||||
     496 CTATTGATAGAAAACATGAAGAAAGCATTAAGACTTATTCAGTCAGAGCT 545

     206 u 206
         |
     546 A 546

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-36

seq_documentation_block:
; Sequence 36, Application US/08916576B
; Patent No. 6171816
;  GENERAL INFORMATION:
;    APPLICANT:  YU, GUO-LIANG
;    APPLICANT:  DILLON, PATRICK J.
;    APPLICANT:  EBNER, REINHARD
;    APPLICANT:  ENDRESS, GREGORY A.
;    TITLE OF INVENTION:  NOVEL HUMAN GROWTH FACTORS
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  US
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/916,576B
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/024,347
;      FILING DATE:  23-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  STEFFE, ERIC K.
;      REGISTRATION NUMBER:  36,688
;      REFERENCE/DOCKET NUMBER:  1488.0500001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  36:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  373 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-916-576B-36

alignment_scores:
            Quality:  487.50              Length:    110
              Ratio:  4.643                Gaps:      3
 Percent Similarity:  95.455   Percent Identity: 91.818

alignment_block:
 US-09-674-266A-181 x US-08-916-576B-36   ..

 Align seg 1/1  to: US-08-916-576B-36  from: 1  to: 373

      16 SerArgArgLeuThrGlnGlyArgTrpValArgLysSerArgValAlaMe 32
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 AGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAGAGTTGCCAT 50

      32 tGluLysIleProValSerAlaPheLeuLeuLeuValAlaLeuSerTyrT 49
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 GGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCTCCTACA 100

      49 hrLeuAlaArgAspThrThrValLysProGlyAlaLysLysAspThrLys 65
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 CTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGACACAAAG 150

      66 AspSerArgProLysLeuProGlnThrLeuSerArgGlyTrpGlyAspGl 82
         |||||||||   ||||||||||||||||||||||||||||||||||||||
     151 GACTCTCGACTAAAACTGCCCCAGACCCTCTCCAGAGGTTGGGGTGACCA 200

      82 nLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerLysThr. 98
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 ACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAATCCAAGACAN 250

      99 SerAsnLysProLeuMetIleIleHisHisLeu.AspGluCysProHisS 115
         :::|||||||||||||||||||||||||||||| ||||||||||||||||
     251 GGCAACAAACCCTTGATGATTATTCATCACTTGGGATGAGTGCCCACACA 300

     115 erGlnAlaLeu...LysLysValPhe 122
         ||::::::::::      |||||||||
     301 GTTCANGTTTTTAANGGAAAGTTTTT 326

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-9

seq_documentation_block:
; Sequence 9, Application US/08916576B
; Patent No. 6171816
;  GENERAL INFORMATION:
;    APPLICANT:  YU, GUO-LIANG
;    APPLICANT:  DILLON, PATRICK J.
;    APPLICANT:  EBNER, REINHARD
;    APPLICANT:  ENDRESS, GREGORY A.
;    TITLE OF INVENTION:  NOVEL HUMAN GROWTH FACTORS
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  US
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/916,576B
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/024,347
;      FILING DATE:  23-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  STEFFE, ERIC K.
;      REGISTRATION NUMBER:  36,688
;      REFERENCE/DOCKET NUMBER:  1488.0500001
```

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  489 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-916-576B-9


alignment_scores:
            Quality:  458.50              Length:    166
              Ratio:  3.728                 Gaps:      9
 Percent Similarity:  74.096   Percent Identity: 70.482

alignment_block:
 US-09-674-266A-181 x US-08-916-576B-9   ..

 Align seg 1/1  to: US-08-916-576B-9  from: 1  to: 489

      11 SerGlyProHisProSerArgArgLeuThrGlnGlyArgTrpValArgLy 27
         |||:::      ||||||||||||||||||||||
       4 TCGGCANAGGATCCTAGCCGCCGACTCACACAAGGC.............. 39

      27 sSerArgValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuV 44
           ||||||||||||||||||||||||||||||||||||||||||||||||
      40 ....AGAGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTG 85

      44 alAlaLeuSerTyrThrLeuAlaArgAspThrThrValLysProGlyAla 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      86 TGGCCCTCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCC 135

      61 LysLysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerAr 77
         ||||||||||||||||||  ||||||||||||||||||||||||||||||
     136 AAAAAGGACACAAAGGACTTTCGACCCAAACTGCCCCAGACCCTCTCCAG 185

      77 gGlyTrpGlyAspGlnLeu.IleTrpThrGlnThrTyrGluGluAla.Le 93
         |||||||||||||||||| |||||||||||||||||||||||||||| ||
     186 AGGTTGGGGTGACCAACTTCATCTGGACTCAGACATATGAAGAAGCTCTT 235

      93 uTyr.LysSerLysThrSerAsnLysProLeuMetIleIleHisHisLeu 109
         |||| ||||||||||||||||||||||||||||||||||||||||||||
     236 ATATNAAATCCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTG 285

     110 AspGluCysProHisSerGlnAla..LeuLysLysValPheAlaGluAsn 125
           ||||||||||||||::::::  ||||||||||||||:::
     286 GGTGAGTGCCCACACAGTTCAAGTTTTTAAAGAAAGTGTTTNGC...... 329

     126 LysGluIleGlnLysLeu.................AlaGluGlnPheVa 136
                         ||||||              :::||||||||||||
     330 .........TGGAAATTAANGAANTCCGGNAATTGGGNNGAGCAGTTTGT 370

     136 l.LeuLeuAsnLeu.ValTyrGluThrThrAspLysHisLeuSerProAs 152
         |       |||||| |||   ||||||      |||||||||
     371 NCCCCNTNAATTTGGGTTTTGGAAACAATGGGCAAACACCTTTTTTGTTG 420

     152 pGlyGlnTyrValProArgIleMetPheValAspPro 164
                      ||||||   :::|||      |||
     421 NTGGNCNATTTTTCCCCGGNTTTTTTTTTTGGCCCC 457

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-10

seq_documentation_block:
; Sequence 10, Application US/08916576B
; Patent No. 6171816
;  GENERAL INFORMATION:
;    APPLICANT:  YU, GUO-LIANG
;    APPLICANT:  DILLON, PATRICK J.
;    APPLICANT:  EBNER, REINHARD
;    APPLICANT:  ENDRESS, GREGORY A.
;    TITLE OF INVENTION:  NOVEL HUMAN GROWTH FACTORS
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  US
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/916,576B
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/024,347
;      FILING DATE:  23-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  STEFFE, ERIC K.
;      REGISTRATION NUMBER:  36,688
;      REFERENCE/DOCKET NUMBER:  1488.0500001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  506 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-916-576B-10


alignment_scores:
            Quality:  376.50              Length:     90
              Ratio:  4.482                 Gaps:      3
 Percent Similarity:  93.333   Percent Identity: 87.778

alignment_block:
 US-09-674-266A-181 x US-08-916-576B-10   ..

 Align seg 1/1  to: US-08-916-576B-10  from: 1  to: 506

      97 LysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCysPr 113
         ||||||||||||||||||||||||||||||||||   ||||||||||||||
      14 AAGACAAGCAACAAACCCTTGATGATTATTCATCANTTGGATGAGTGCCC 63

     113 oHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGlnL 130
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      64 ACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCAGA 113

     130 ysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThr.As 146
         |||||||||||||||||||||||||||||||||||||||||||||||| ||
     114 AATTGGCAGAGCAGTTGTCCTCCTCAATCTGGTTTATGAAACAACTGGA 163

     146 pLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheValA 163
         |||||||||||||||||||||||||||||||||||:::||||||||||
     164 CAAACACCTTTCTCCTGATGGCCAGTATGTNCCCAGGGTTATGTTTGTTG 213

     163 spProSerLeu.ThrValArgAlaAspIleThrGlyArg...TyrSerAs 178
         |||||:::||| ||||||||||||||||||||||||||   ::::::::
     214 ACCCATNTCTGGACAGTTAGAGCCGATATCACTGGGAGGATATTTCAAAC 263

     178 nArgLeuTyrAlaTyr 183
         :      |||||||||
     264 CGTNNCTTATGCTTAC 279

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-37
```

```
seq_documentation_block:
; Sequence 37, Application US/08916576B
; Patent No. 6171816
;  GENERAL INFORMATION:
;    APPLICANT:  YU, GUO-LIANG
;    APPLICANT:  DILLON, PATRICK J.
;    APPLICANT:  EBNER, REINHARD
;    APPLICANT:  ENDRESS, GREGORY A.
;    TITLE OF INVENTION:  NOVEL HUMAN GROWTH FACTORS
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  US
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/916,576B
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/024,347
;      FILING DATE:  23-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  STEFFE, ERIC K.
;      REGISTRATION NUMBER:  36,688
;      REFERENCE/DOCKET NUMBER:  1488.0500001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  37:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  158 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-916-576B-37


alignment_scores:
            Quality:  253.00             Length:     51
              Ratio:  4.961                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x US-08-916-576B-37   ..

 Align seg 1/1  to: US-08-916-576B-37  from: 1  to: 158

      15 ProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerArgValAl 31
         |||||||||||||||||||||||||||||||||||||||||||||||||||
       6 CCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAGAGTTGC 55

      31 aMetGluLysIleProValSerAlaPheLeuLeuLeuValAlaLeuSerT 48
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      56 CATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCTCCT 105

      48 yrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLysAspThr 64
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     106 ACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGACACA 155

      65 Lys 65
         |||
     156 AAG 158
```

```
seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-030-607-209

seq_documentation_block:
; Sequence 209, Application US/09030607
; Patent No. 6262245
;  GENERAL INFORMATION:
;    APPLICANT:  Xu, Jiangchun .
;    APPLICANT:  Dillon, Davin C.
;    TITLE OF INVENTION:  COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS
;    NUMBER OF SEQUENCES:  224
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/030,607
;      FILING DATE:  25-FEB-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.427C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  209:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  159 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-030-607-209


alignment_scores:
            Quality:  252.00             Length:     52
              Ratio:  4.846                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x US-09-030-607-209   ..

 Align seg 1/1  to: US-09-030-607-209  from: 1  to: 159

      25 ValArgLysSerArgValAlaMetGluLysIleProValSerAlaPheLe 41
         |||||||||||||||||||||||||||||||||||||||||||||||||||
       3 GTGAGGAAATCCAGAGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTT 52

      41 uLeuLeuValAlaLeuSerTyrThrLeuAlaArgAspThrThrValLysP 58
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      53 GCTCCTTGTGGCCCTCTCCTACACTCTGGCCAGAGATACCACAGTCAAAC 102

      58 roGlyAlaLysLysAspThrLysAspSerArgProLysLeuProGlnThr 74
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     103 CTGGAGCCAAAAAGGACACAAAGGACTCTCGACCCAAACTGCCCCAGACC 152

      75 LeuSer 76
         ||||||
     153 CTCTCC 158

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-439-313-209

seq_documentation_block:
; Sequence 209, Application US/09439313
```

```
; Patent No. 6329505
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Mitcham, Jennifer L.
;  APPLICANT:  Harlocker, Susan Louise
;  APPLICANT:  Jiang Yuqui
;  APPLICANT:  Reed, Steven G.
;  APPLICANT:  Kalos, Michael
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Retter, Mark
;  APPLICANT:  Solk, John
;  APPLICANT:  Day, Craig
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF PROSTATE CANCER
;  FILE REFERENCE: 210121.427C9
;  CURRENT APPLICATION NUMBER: US/09/439,313
;  CURRENT FILING DATE:  1999-11-12
;  NUMBER OF SEQ ID NOS: 575
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 209
;   LENGTH: 159
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-439-313-209


alignment_scores:
            Quality:  252.00              Length:     52
              Ratio:  4.846                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x US-09-439-313-209   ..

 Align seg 1/1  to: US-09-439-313-209  from: 1  to: 159

      25 ValArgLysSerArgValAlaMetGluLysIleProValSerAlaPheLe 41
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       3 GTGAGGAAATCCAGAGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTT 52

      41 uLeuLeuValAlaLeuSerTyrThrLeuAlaArgAspThrThrValLysP 58
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      53 GCTCCTTGTGGCCCTCTCCTACACTCTGGCCAGAGATACCACAGTCAAAC 102

      58 roGlyAlaLysLysAspThrLysAspSerArgProLysLeuProGlnThr 74
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     103 CTGGAGCCAAAAAGGACACAAAGGACTCTCGACCCAAACTGCCCCAGACC 152

      75 LeuSer 76
         ||||||
     153 CTCTCC 158

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-3

seq_documentation_block:
; Sequence 3, Application US/08916576B
; Patent No. 6171816
;  GENERAL INFORMATION:
;    APPLICANT:  YU, GUO-LIANG
;    APPLICANT:  DILLON, PATRICK J.
;    APPLICANT:  EBNER, REINHARD
;    APPLICANT:  ENDRESS, GREGORY A.
;    TITLE OF INVENTION:  NOVEL HUMAN GROWTH FACTORS
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  US
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/916,576B
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/024,347
;      FILING DATE:  23-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  STEFFE, ERIC K.
;      REGISTRATION NUMBER:  36,688
;      REFERENCE/DOCKET NUMBER:  1488.0500001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1423 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  88..603
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  157..603
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  88..156
US-08-916-576B-3


alignment_scores:
            Quality:  222.00              Length:    158
              Ratio:  2.242                Gaps:      3
 Percent Similarity:  62.658   Percent Identity: 34.177

alignment_block:
 US-09-674-266A-181 x US-08-916-576B-3   ..

 Align seg 1/1  to: US-08-916-576B-3  from: 1  to: 1423

      66 AspSerArgProLysLeuProGlnThr..................... 74
         ::::::||||||:::|||      |||
      91 GAGACGCGGCCTCGTCTCGGGGCCACCTGTTTGCTGGGCTTCAGTTTCCT 140

      75 ..............................LeuSerArgGlyTrpG 80
                                       |||:::::::|||:::|
     141 GCTCCTCGTCATCTCTTCTGATGGACATAATGGGCTTGGAAAGGGTTTTG 190

      80 lyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysSer 96
         |||||::::::   |||   :::|||   |||::::::      ::::::
     191 GAGATCATATTCATTGG...AGGACACTGGAAGATGGGAAGAAAGAAGCA 237

      97 LysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCysPr 113
            :::|||:::   ||||||||||:::|||   |||   :::   |||
     238 GCTGCCAGTGGACTGCCCCTGATGGTGATTATTCATAAATCCTGGTGTGG 287

     113 oHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGlnL 130
               :::||||||||||      |||||||||:::   ||||||:::: 
     288 AGCTTGCAAAGCTCTAAAGCCCAAATTTGCAGAATCTACGGAAATTTCAG 337

     130 ysLeuAlaGluGlnPheValLeuLeuAsnLeu......ValTyrGluThr 144
         ::|||::::::::::||||||::::::||||||             |||
     338 AACTCTCCCATAATTTTGTTATGGTAAATCTTGAGGATGAAGAGGAACCC 387

     145 ThrAspLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPh 161
            |||:::   :::||||||||||||   |||:::||||||||||:::||
```

```
388 AAAGATGAAGATTTCAGCCCTGACGGGGGTTATATTCCACGAATCCTTTT 437

161 eValAspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerA 178
    |:::|||||||||      |||:::   :::|||   ::::::   :::|
438 TCTGGATCCCAGTGGCAAGGTGCATCCTGAAATCATCAATGAGAATGGAA 487

178 snArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsn 194
    ||      |||   |||      :::::::::      ::::::::::::
488 ACCCCAGCTACAAGTATTTTTATGTCAGTGCCGAGCAAGTTGTTCAGGGG 537

195 MetLysLysAlaLeuLysLeuLeu 202
    ||||||:::|||   :::   |||
538 ATGAAGGAAGCTCAGGAAAGGCTG 561
```

```
OM of: US-09-674-266A-181 to: EST:*   out_format : pfs

Date: Oct 8, 2002 10:02 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlh
 -Q=/cgn2_1/USPTO_spool/US09674266/runat_08102002_093323_26150/app_query.fasta_1.271
 -DB=EST -QFMT=fastap -SUFFIX=p2n.rst -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500
 -MINLEN=0 -MAXLEN=2000000000 -USER=US09674266_@CGN1_1_2564
 -NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-674-266A-181
Query length: 206
Database: EST:*
Database sequences: 13736207
Database length: -1841457050
Search time (sec): 1720.980000


score_list:
Sequence          Strd Orig    ZScore   EScore Len  ! Documentation  ..
gb_est2:BG778248    + 1008.00 1997.61  5.1e-102  666  ! BG778248 602666830F1 NIH_MGC_60
gb_est1:AW582256    + 1008.00 1997.25  5.3e-102  689  ! AW582256 QV4-ST0212-120100-075-
gb_est2:BI523582    + 1002.00 1986.32  2.2e-101  624  ! BI523582 603051712F1 NIH_MGC_12
gb_est2:BG179160    +  983.00 1946.37  3.6e-99   750  ! BG179160 602330032F1 NIH_MGC_91
gb_est2:BE748141    +  980.00 1941.28  7.0e-99   689  ! BE748141 601571509F1 NIH_MGC_55
gb_est2:BF733142    +  980.00 1940.30  7.9e-99   755  ! BF733142 EST039 Human hepatocel
gb_est2:BG501149    +  979.00 1938.52  9.9e-99   740  ! BG501149 602546389F1 NIH_MGC_60
gb_est2:BI770077    +  976.00 1931.85  2.3e-98   788  ! BI770077 603053412F1 NIH_MGC_12
gb_est1:AW956284    +  975.00 1930.81  2.7e-98   721  ! AW956284 EST368354 MAGE reseque
gb_est2:BG285246    +  971.50 1921.85  8.4e-98   866  ! BG285246 602409569F1 NIH_MGC_91
gb_est2:BM129777    +  971.00 1925.12  5.5e-98   581  ! BM129777 if23a08.y1 Melton Norm
gb_est2:BE796856    +  962.00 1901.89  1.1e-96   948  ! BE796856 601588166F1 NIH_MGC_7
gb_est2:BG610804    +  961.00 1901.02  1.2e-96   853  ! BG610804 602612442F1 NIH_MGC_60
gb_est2:BE870718    +  957.00 1894.64  2.8e-96   734  ! BE870718 601448570F1 NIH_MGC_65
gb_est2:BM006356    +  948.00 1876.31  2.9e-95   758  ! BM006356 603615055F1 NIH_MGC_11
gb_est2:BM006510    +  948.00 1875.77  3.1e-95   797  ! BM006510 603615461F1 NIH_MGC_11
gb_est2:BM006784    +  948.00 1875.70  3.1e-95   802  ! BM006784 603615217F1 NIH_MGC_11
gb_est2:BM006554    +  948.00 1875.13  3.4e-95   846  ! BM006554 603615521F1 NIH_MGC_11
gb_est2:BM006458    +  948.00 1874.92  3.5e-95   863  ! BM006458 603615186F1 NIH_MGC_11
gb_est2:BM006669    +  948.00 1874.59  3.6e-95   890  ! BM006669 603615373F1 NIH_MGC_11
gb_est2:BI763228    +  944.00 1870.76  5.9e-95   603  ! BI763228 603049976F1 NIH_MGC_11
gb_est1:AA316115    +  943.00 1869.44  7.0e-95   566  ! AA316115 EST187838 Colon carcin
gb_est2:BG823844    +  942.00 1864.54  1.3e-94   742  ! BG823844 602729503F1 NIH_MGC_15
gb_est2:BG386151    +  940.00 1860.42  2.2e-94   751  ! BG386151 602455333F1 NIH_MGC_15
gb_est2:BF680376    +  931.50 1842.06  2.3e-93   853  ! BF680376 602154161F1 NIH_MGC_83
gb_est1:AL543432    -  925.00 1830.00  1.1e-92   782  ! AL543432 AL543432 LTI_NFL006_PL
gb_est1:AA307697    +  919.00 1821.76  3.2e-92   551  ! AA307697 EST178713 Colon carcin
gb_est1:AA314225    +  899.00 1782.06  5.1e-90   537  ! AA314225 EST186091 Colon carcin
gb_est2:BF680021    +  895.00 1770.53  2.3e-89   747  ! BF680021 602154740F1 NIH_MGC_83
gb_est2:BF207999    +  892.50 1763.79  5.4e-89   879  ! BF207999 601862323F1 NIH_MGC_53
gb_est2:BE867131    +  889.00 1760.49  8.2e-89   623  ! BE867131 601442973F1 NIH_MGC_65
gb_est2:BI770277    +  886.00 1752.07  2.4e-88   781  ! BI770277 603055968F1 NIH_MGC_12
gb_est2:BI760685    +  879.00 1736.57  1.8e-87   899  ! BI760685 603044871F1 NIH_MGC_11
gb_est2:BE737345    +  878.50 1738.66  1.3e-87   674  ! BE737345 601306577F1 NIH_MGC_39
gb_est1:AW369331    +  872.00 1726.44  6.5e-87   627  ! AW369331 QV1-BN0002-061299-039-
gb_est2:BE736553    +  862.00 1702.30  1.4e-85   924  ! BE736553 601306131F1 NIH_MGC_39
gb_est1:AW369360    +  861.00 1703.94  1.2e-85   658  ! AW369360 QV1-BN0002-061299-039-
gb_est2:BG528540    +  851.00 1680.49  2.3e-84   909  ! BG528540 602580027F1 NIH_MGC_60
gb_est2:BE565542    +  848.50 1676.77  3.8e-84   807  ! BE565542 601338310F1 NIH_MGC_53
gb_est1:AL543472    +  843.00 1671.06  7.8e-84   493  ! AL543472 AL543472 LTI_NFL006_PL
gb_est2:BG962742    +  842.00 1665.27  1.6e-83   702  ! BG962742 602830265F1 NCI_CGA
gb_htc:AK007677     +  842.00 1664.17  1.9e-83   778  ! AK007677 Mus musculus 10 day
gb_est2:BG964303    +  836.00 1654.19  6.8e-83   645  ! BG964303 602831911F1 NCI_CGA
gb_est2:BG172599    +  835.50 1648.87  1.3e-82   965  ! BG172599 602337015F1 NCI_CGA
gb_est2:BF801670    -  835.00 1653.11  7.8e-83   592  ! BF801670 PM4-CI0079-281000-0

seq_name: gb_est2:BG778248

seq_documentation_block:
LOCUS       BG778248               666 bp    mRNA    linear   EST 15-MAY-2001
DEFINITION  602666830F1 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:4806679 5',
            mRNA sequence.
ACCESSION   BG778248
VERSION     BG778248.1  GI:14048565
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 666)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1656  row: 1  column: 08
            High quality sequence stop: 665.
FEATURES             Location/Qualifiers
     source          1. .666
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4806679"
                     /clone_lib="NIH_MGC_60"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: prostate; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                     ); Double-stranded cDNA was prepared from cell line RNA.
                     5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).  Average
                     insert size 1.5 kb (range 0.9-4.0 kb). 14/15 colonies
                     contained inserts  by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
                     Library."
BASE COUNT      205 a    165 c    140 g    156 t
ORIGIN


alignment_scores:
            Quality: 1008.00             Length:    195
              Ratio:  5.169                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x BG778248   ..

 Align seg 1/1  to: BG778248  from: 1  to: 666

      12 GlyProHisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSe 28
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      12 GGACCGCATCCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATC 61

      28 rArgValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValA 45
```

```
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  62 CAGAGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGG 111

  45 laLeuSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLys 61
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 112 CCCTCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAA 161

  62 LysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGl 78
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 162 AAGGACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGG 211

  78 yTrpGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrL 95
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 212 TTGGGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATA 261

  95 ysSerLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGlu 111
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 262 AATCCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAG 311

 112 CysProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIl 128
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 312 TGCCCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAAT 361

 128 eGlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrT 145
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 362 CCAGAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAA 411

 145 hrAspLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPhe 161
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 412 CTGACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTT 461

 162 ValAspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAs 178
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 462 GTTGACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAA 511

 178 nArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnM 195
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 512 TCGTCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACA 561

 195 etLysLysAlaLeuLysLeuLeuLysThrGluLeu 206
     ||||||||||||||||||||||||||||||||||
 562 TGAAGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 596

seq_name: gb_est1:AW582256

seq_documentation_block:
LOCUS       AW582256                689 bp    mRNA    linear   EST 16-MAR-2000
DEFINITION  QV4-ST0212-120100-075-e10 ST0212 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW582256
VERSION     AW582256.1  GI:7257305
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 689)
  AUTHORS   HCGP http://www.ludwig.org.br/ORESTES.
  TITLE     The FAPESP/LICR Human Cancer Genome Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV4&t2=QV4-ST0212-
            120100-075-e10&t3=2000-01-12&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 687.

FEATURES             Location/Qualifiers
     source          1. .689
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="ST0212"
                     /dev_stage="Adult"
                     /note="Organ: stomach; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      212 a    172 c    143 g    162 t
ORIGIN


alignment_scores:
            Quality: 1008.00             Length:    195
              Ratio:  5.169                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x AW582256   ..

 Align seg 1/1  to: AW582256  from: 1  to: 689

  12 GlyProHisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSe 28
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  25 GGACCGCATCCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATC 74

  28 rArgValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValA 45
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  75 CAGAGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGG 124

  45 laLeuSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLys 61
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 125 CCCTCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAA 174

  62 LysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGl 78
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 175 AAGGACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGG 224

  78 yTrpGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrL 95
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 225 TTGGGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATA 274

  95 ysSerLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGlu 111
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 275 AATCCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAG 324

 112 CysProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIl 128
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 325 TGCCCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAAT 374

 128 eGlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrT 145
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 375 CCAGAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAA 424

 145 hrAspLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPhe 161
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 425 CTGACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTT 474

 162 ValAspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAs 178
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 475 GTTGACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAA 524

 178 nArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnM 195
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 525 TCGTCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACA 574

 195 etLysLysAlaLeuLysLeuLeuLysThrGluLeu 206
     ||||||||||||||||||||||||||||||||||
```

```
    575 TGAAGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 609

seq_name: gb_est2:BI523582

seq_documentation_block:
LOCUS       BI523582               624 bp    mRNA    linear   EST 29-AUG-2001
DEFINITION  603051712F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5201331 5',
            mRNA sequence.
ACCESSION   BI523582
VERSION     BI523582.1  GI:15348374
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 624)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11504  row: h  column: 04
            High quality sequence start: 5
            High quality sequence stop: 624.
FEATURES             Location/Qualifiers
     source          1. .624
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5201331"
                     /clone_lib="NIH_MGC_122"
                     /lab_host="DH10B"
                     /note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
                     Site_1: NotI; Site_2: EcoRV (destroyed); RNA source
                     anonymous pool of 24 week female lung, 16 week female
                     spleen, and 20-22 week male spleens.  Library is oligo-dT
                     primed and directionally cloned (EcoRV site is destroyed
                     upon  cloning).  Average insert size 1.4 kb, insert size
                     range 1-3 kb. Library is normalized and enriched for
                     full-length clones and was  constructed by C. Gruber
                     (Invitrogen).  Research Genetics tracking  code 026. Note:
                     this is a NIH_MGC Library."
BASE COUNT      190 a    161 c    126 g    147 t
ORIGIN


alignment_scores:
           Quality: 1002.00             Length:    194
             Ratio:  5.165                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x BI523582   ..

 Align seg 1/1  to: BI523582  from: 1  to: 624

      13 ProHisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerAr 29
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       2 CCGCATCCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAG 51

      29 gValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAlaL 46
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      52 AGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCC 101

      46 euSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLys 62
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     102 TCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAG 151

      63 AspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTr 79
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     152 GACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTG 201

      79 pGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysS 96
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     202 GGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAAT 251

      96 erLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCys 112
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     252 CCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGC 301

     113 ProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGl 129
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     302 CCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCA 351

     129 nLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrA 146
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     352 GAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTG 401

     146 spLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheVal 162
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     402 ACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTT 451

     163 AspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnAr 179
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     452 GACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAATCG 501

     179 gLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetL 196
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     502 TCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACATGA 551

     196 ysLysAlaLeuLysLeuLeuLysThrGluLeu 206
         |||||||||||||||||||||||||||||||
     552 AGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 583

seq_name: gb_est2:BG179160

seq_documentation_block:
LOCUS       BG179160               750 bp    mRNA    linear   EST 06-FEB-2001
DEFINITION  602330032F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4431583 5',
            mRNA sequence.
ACCESSION   BG179160
VERSION     BG179160.1  GI:12685943
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 750)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10186  row: o  column: 08
            High quality sequence stop: 732.
FEATURES             Location/Qualifiers
     source          1. .750
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4431583"
                     /clone_lib="NIH_MGC_91"
                     /tissue_type="adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
```

```
                  /note="Organ: prostate; Vector: pCMV-SPORT6; Site_1: NotI;
                  Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                  Average insert size 1.4 kb. Library enriched for
                  full-length clones and constructed by Life Technologies.
                  Note: this is a NIH_MGC Library."
BASE COUNT      234 a    174 c    158 g    183 t      1 others
ORIGIN


alignment_scores:
            Quality:  983.00             Length:    191
              Ratio:  5.147                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x BG179160   ..

 Align seg 1/1  to: BG179160  from: 1  to: 750

      15 ProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerArgValAl 31
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      15 CCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAGAGTTGC 64

      31 aMetGluLysIleProValSerAlaPheLeuLeuLeuValAlaLeuSerT 48
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      65 CATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCTCCT 114

      48 yrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLysAspThr 64
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     115 ACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGACACA 164

      65 LysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTrpGlyAs 81
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     165 AAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGGGTGA 214

      81 pGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerLysT 98
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     215 CCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAATCCAAGA 264

      98 hrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCysProHis 114
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     265 CAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGCCCACAC 314

     115 SerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGlnLysLe 131
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     315 AGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCAGAAATT 364

     131 uAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrAspLysH 148
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     365 GGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTGACAAAC 414

     148 isLeuSerProAspGlyGlnTyrValProArgIleMetPheValAspPro 164
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     415 ACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTTGACCCA 464

     165 SerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnArgLeuTy 181
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     465 TCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAATCGTCTCTA 514

     181 rAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLysA 198
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     515 TGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACATGAAGAAAG 564

     198 laLeuLysLeuLeuLysThrGlu 205
         ||||||||||||||||||||||
     565 CTCTCAAGTTGCTGAAGACTGAA 587

seq_name: gb_est2:BE748141

seq_documentation_block:
LOCUS       BE748141               689 bp    mRNA    linear   EST 15-SEP-2000
DEFINITION  601571509F1 NIH_MGC_55 Homo sapiens cDNA clone IMAGE:3838347 5',
            mRNA sequence.
ACCESSION   BE748141
VERSION     BE748141.1  GI:10162133
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 689)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM525  row: a  column: 04
            High quality sequence stop: 643.
FEATURES             Location/Qualifiers
     source          1. .689
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3838347"
                     /clone_lib="NIH_MGC_55"
                     /tissue_type="from acute myelogenous leukemia"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: bone marrow; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                     ); Double-stranded cDNA was prepared from cell line RNA.
                     5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).   Average
                     insert size 1.65 kb (range 0.9-4.0 kb).  14/15 colonies
                     contained  inserts by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT      209 a    170 c    148 g    162 t
ORIGIN


alignment_scores:
            Quality:  980.00             Length:    191
              Ratio:  5.131                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 99.476

alignment_block:
 US-09-674-266A-181 x BE748141   ..

 Align seg 1/1  to: BE748141  from: 1  to: 689

      15 ProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerArgValAl 31
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
       3 CCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAGAGTTGC 52

      31 aMetGluLysIleProValSerAlaPheLeuLeuLeuValAlaLeuSerT 48
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      53 CATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCTCCT 102

      48 yrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLysAspThr 64
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     103 ACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGACACA 152

      65 LysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTrpGlyAs 81
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     153 AAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGGGTGA 202

      81 pGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerLysT 98
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
     203 CCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAATCCAAGA 252

      98 hrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCysProHis 114
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     253 CAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGCCCACAC 302

     115 SerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGlnLysLe 131
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     303 AGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCAGAAATT 352

     131 uAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrAspLysH 148
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     353 GGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTGACAAAC 402

     148 isLeuSerProAspGlyGlnTyrValProArgIleMetPheValAspPro 164
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     403 ACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTTGACCCA 452

     165 SerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnArgLeuTy 181
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     453 TCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAACCGTCTCTA 502

     181 rAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLysA 198
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     503 TGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACATGAAGAAAG 552

     198 laLeuLysLeuLeuLysThrGlu 205
         ||||||||||||||||||:::
     553 CTCTCAAGTTGCTGAAGACTGAC 575

seq_name: gb_est2:BF733142

seq_documentation_block:
LOCUS       BF733142               755 bp    mRNA    linear   EST 09-JAN-2001
DEFINITION  EST039 Human hepatocellular carcinoma subtracted cDNA library Homo
            sapiens cDNA clone P39 5', mRNA sequence.
ACCESSION   BF733142
VERSION     BF733142.1  GI:12058378
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 755)
  AUTHORS   Zhu,W., Duan,F., Liu,D., Ma,J., Bai,J. and Gao,T.
  TITLE     Suppression subtracted hybridization to identify differentially
            expressed genes of hepatocellular carcinoma and expressed sequence
            tags sequencing
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Wuling Zhu
            Department of Nucleic Acid Research
            Institute of Digestive Disease
            2 Jingba Road, Zhengzhou, 450003, Henan Province, PR.China
            Tel: 86 0371 3921444
            Fax: 86 0371 6960571
            Email: wuling_z@hotmail.com
            Human hepatocellular carcinoma cDNA research supported by Institute
            of Digestive Disease, Henan Medical University; cDNA insert
            sequencing: Genetech Biotechnology Company Limited. cDNA library
            construction: Department of Nucleic Acid Research, Institute of
            Digestive Disease.
            Seq primer: T7.
FEATURES             Location/Qualifiers
     source          1. .755
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="P39"
                     /clone_lib="Human hepatocellular carcinoma subtracted cDNA
                     library"
                     /note="Organ: Liver"
BASE COUNT      237 a    171 c    158 g    189 t
ORIGIN
```

```
alignment_scores:
            Quality:  980.00              Length:    192
              Ratio:  5.131                 Gaps:      0
 Percent Similarity:  99.479   Percent Identity: 99.479

alignment_block:
 US-09-674-266A-181 x BF733142   ..

 Align seg 1/1  to: BF733142  from: 1  to: 755

      15 ProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerArgValAl 31
         |||||||||||||||||||||||||||||||||||||||||||||||||||
       3 CCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAGAGTTGC 52

      31 aMetGluLysIleProValSerAlaPheLeuLeuLeuValAlaLeuSerT 48
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      53 CATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCTCCT 102

      48 yrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLysAspThr 64
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     103 ACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGACACA 152

      65 LysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTrpGlyAs 81
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     153 AAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGGGTGA 202

      81 pGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerLysT 98
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     203 CCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAATCCAAGA 252

      98 hrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCysProHis 114
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     253 CAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGCCCACAC 302

     115 SerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGlnLysLe 131
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     303 AGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCAGAAATT 352

     131 uAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrAspLysH 148
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     353 GGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTGACAAAC 402

     148 isLeuSerProAspGlyGlnTyrValProArgIleMetPheValAspPro 164
         ||||||||||||   ||||||||||||||||||||||||||||||||||||
     403 ACCTTTCTCCTGGTGGCCAGTATGTCCCCAGGATTATGTTTGTTGACCCA 452

     165 SerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnArgLeuTy 181
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     453 TCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAACCGTCTCTA 502

     181 rAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLysA 198
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     503 TGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACATGAAGAAAG 552

     198 laLeuLysLeuLeuLysThrGluLeu 206
         |||||||||||||||||||||||||
     553 CTCTCAAGTTGCTGAAGACTGAATTG 578

seq_name: gb_est2:BG501149

seq_documentation_block:
LOCUS       BG501149               740 bp    mRNA    linear   EST 27-MAR-2001
DEFINITION  602546389F1 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:4668575 5',
            mRNA sequence.
ACCESSION   BG501149
VERSION     BG501149.1  GI:13462666
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 740)
```

```
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1479  row: a  column: 24
            High quality sequence stop: 738.
FEATURES             Location/Qualifiers
     source          1. .740
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4668575"
                     /clone_lib="NIH_MGC_60"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: prostate; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                     ); Double-stranded cDNA was prepared from cell line RNA.
                     5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).  Average
                     insert size 1.5 kb (range 0.9-4.0 kb). 14/15 colonies
                     contained inserts  by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
                     Library."
BASE COUNT      233 a    171 c    157 g    179 t
ORIGIN


alignment_scores:
            Quality:  979.00             Length:    195
              Ratio:  5.021                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 99.487

alignment_block:
 US-09-674-266A-181 x BG501149   ..

 Align seg 1/1  to: BG501149  from: 1  to: 740

      12 GlyProHisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSe 28
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      14 GGACCGCATCCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATC 63

      28 rArgValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValA 45
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      64 CAGAGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGG 113

      45 laLeuSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLys 61
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     114 CCCTCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAA 163

      62 LysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGl 78
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     164 AAGGACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGG 213

      78 yTrpGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrL 95
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     214 TTGGGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATA 263

      95 ysSerLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGlu 111
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     264 AATCCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAG 313

     112 CysProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIl 128
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     314 TGCCCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAAT 363

     128 eGlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrT 145
         ||||||||||||||||||   ||||||||||||||||||||||||||||
     364 CCAGAAATTGGCAGAGCAGTT.GTCCTCCTCAATCTGGTTTATGAAACAA 412

     145 hrAspLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPhe 161
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     413 CTGACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTT 462

     162 ValAspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAs 178
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     463 GTTGACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAA 512

     178 nArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnM 195
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     513 TCGTCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACA 562

     195 etLysLysAlaLeuLysLeuLeuLysThrGluLeu 206
         |||||||||||||||||||||||||||||||||||
     563 TGAAGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 597

seq_name: gb_est2:BI770077

seq_documentation_block:
LOCUS       BI770077                 788 bp    mRNA    linear   EST 25-SEP-2001
DEFINITION  603053412F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5202866 5',
            mRNA sequence.
ACCESSION   BI770077
VERSION     BI770077.1  GI:15761655
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 788)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11508  row: h  column: 03
            High quality sequence stop: 786.
FEATURES             Location/Qualifiers
     source          1. .788
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5202866"
                     /clone_lib="NIH_MGC_122"
                     /lab_host="DH10B"
                     /note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
                     Site_1: NotI; Site_2: EcoRV (destroyed); RNA source
                     anonymous pool of 24 week female lung, 16 week female
                     spleen, and 20-22 week male spleens.  Library is oligo-dT
                     primed and directionally cloned (EcoRV site is destroyed
                     upon  cloning).  Average insert size 1.4 kb, insert size
                     range 1-3 kb. Library is normalized and enriched for
                     full-length clones and was  constructed by C. Gruber
                     (Invitrogen).  Research Genetics tracking  code 026. Note:
                     this is a NIH_MGC Library."
BASE COUNT      248 a    178 c    165 g    197 t
ORIGIN


alignment_scores:
```

```
          Quality:  976.00            Length:    194
            Ratio:  5.083               Gaps:      1
 Percent Similarity:  98.969   Percent Identity: 98.969

alignment_block:
 US-09-674-266A-181 x BI770077   ..

 Align seg 1/1  to: BI770077  from: 1  to: 788

      14 HisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerArgVa 30
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
       3 CATCCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAGAGT 52

      30 lAlaMetGlu.LysIleProValSerAlaPheLeuLeuLeuValAlaLeu 46
         |||||||||| ||||||||||||||||||||||||||||||||||||||||
      53 TGCCATGGAGGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTC 102

      47 SerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLysAs 63
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     103 TCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGA 152

      63 pThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTrpG 80
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     153 CACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGG 202

      80 lyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysSer 96
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     203 GTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAATCC 252

      97 LysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCysPr 113
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     253 AAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGCCC 302

     113 oHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGlnL 130
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     303 ACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCAGA 352

     130 ysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrAsp 146
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     353 AATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTGAC 402

     147 LysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheValAs 163
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     403 AAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTTGA 452

     163 pProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnArgL 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     453 CCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAATCGTC 502

     180 euTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetLys 196
         |||||||||||||||||||||||||||||||   ||||||||||||||||
     503 TCTATGCTTACGAACCTGCAGATACAGCTCTGTGGCTTGACAACATGAAG 552

     197 LysAlaLeuLysLeuLeuLysThrGluLeu 206
         |||||||||||||||||||||||||||||
     553 AAAGCTCTCAAGTTGCTGAAGACTGAATTG 582

seq_name: gb_est1:AW956284

seq_documentation_block:
LOCUS       AW956284                721 bp    mRNA    linear   EST 01-JUN-2000
DEFINITION  EST368354 MAGE resequences, MAGD Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW956284
VERSION     AW956284.1  GI:8145967
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 721)
  AUTHORS   Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
            ,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
            Quackenbush,J.
  TITLE     Assessment of gene expression patterns in a model of colon tumor
            metastasis using a 19,200 element cDNA microarray
  JOURNAL   Unpublished (2000)
COMMENT     Contact: John Quackenbush
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 3528
            Fax: 301 838 0208
            Email: johnq@tigr.org
            Plate: 90
            Seq primer: Reverse.
FEATURES             Location/Qualifiers
     source          1. .721
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="MAGE resequences, MAGD"
                     /note="Vector: pBluescriptSKm"
BASE COUNT      224 a    173 c    148 g    175 t      1 others
ORIGIN


alignment_scores:
          Quality:  975.00            Length:    194
            Ratio:  5.078               Gaps:      0
 Percent Similarity:  98.969   Percent Identity: 97.938

alignment_block:
 US-09-674-266A-181 x AW956284   ..

 Align seg 1/1  to: AW956284  from: 1  to: 721

      13 ProHisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerAr 29
         |||      ::::::||||||||||||||||||||||||||||||||||||
       6 CCGACTCGTGCCGAACGACTCACACAAGGCAGGTGGGTGAGGAAATCCAG 55

      29 gValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAlaL 46
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      56 AGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCC 105

      46 euSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLys 62
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     106 TCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAG 155

      63 AspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTr 79
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     156 GACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTG 205

      79 pGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysS 96
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     206 GGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAAT 255

      96 erLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCys 112
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     256 CCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGC 305

     113 ProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGl 129
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     306 CCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCA 355

     129 nLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrA 146
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     356 GAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTG 405

     146 spLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheVal 162
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     406 ACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTT 455

     163 AspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnAr 179
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     456 GACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAATCG 505

     179 gLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetL 196
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
     506 TCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACATGA 555

     196 ysLysAlaLeuLysLeuLeuLysThrGluLeu 206
         ||||||||||||||||||||||||||||||
     556 AGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 587

seq_name: gb_est2:BG285246

seq_documentation_block:
LOCUS       BG285246                866 bp    mRNA    linear   EST 21-FEB-2001
DEFINITION  602409569F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4539353 5',
            mRNA sequence.
ACCESSION   BG285246
VERSION     BG285246.1  GI:13037011
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 866)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10467  row: i  column: 18
            High quality sequence stop: 750.
FEATURES             Location/Qualifiers
     source          1. .866
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4539353"
                     /clone_lib="NIH_MGC_91"
                     /tissue_type="adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: prostate; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.4 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      270 a    194 c    189 g    212 t      1 others
ORIGIN


alignment_scores:
            Quality:  971.50             Length:    194
              Ratio:  5.086                Gaps:      1
 Percent Similarity:  98.454   Percent Identity: 98.454

alignment_block:
 US-09-674-266A-181 x BG285246   ..

 Align seg 1/1  to: BG285246  from: 1  to: 866

      14 HisProSer...ArgArgLeuThrGlnGlyArgTrpValArgLysSerAr 29
         |||   |||   ||||||||||||||||||||||||||||||||||||||
      18 CACGCGTCCGGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAG 67

      29 gValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAlaL 46
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      68 AGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCC 117

      46 euSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLys 62
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     118 TCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAG 167

      63 AspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTr 79
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     168 GACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTG 217

      79 pGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysS 96
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     218 GGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAAT 267

      96 erLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCys 112
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     268 CCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGC 317

     113 ProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGl 129
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     318 CCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCA 367

     129 nLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrA 146
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     368 GAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTG 417

     146 spLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheVal 162
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     418 ACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTT 467

     163 AspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnAr 179
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     468 GACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAATCG 517

     179 gLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetL 196
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     518 TCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACATGA 567

     196 ysLysAlaLeuLysLeuLeuLysThrGluLeu 206
         ||||||||||||||   ||||||||||||||
     568 AGAAAGCTCTCAAGTNGCTGAAGACTGAATTG 599

seq_name: gb_est2:BM129777

seq_documentation_block:
LOCUS       BM129777                581 bp    mRNA    linear   EST 27-NOV-2001
DEFINITION  if23a08.y1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
            cDNA 5' similar to TR:O95994 O95994 SECRETED CEMENT GLAND PROTEIN
            XAG-2 HOMOLOG. ;, mRNA sequence.
ACCESSION   BM129777
VERSION     BM129777.1  GI:17124329
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 581)
  AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Sceance,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
            ,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.
            , Jackson,Y. and Bowers,Y.
  TITLE     Endocrine Pancreas Consortium
  JOURNAL   Unpublished (2000)
COMMENT     Other_ESTs: if23a08.x1
            Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Douglas Melton DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Juliana Brown
            (brown@fas.harvard.edu)
            High quality sequence stop: 429.
```

```
FEATURES             Location/Qualifiers
     source          1. .581
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
                     /sex="Both"
                     /tissue_type="Islets of Langerhans"
                     /dev_stage="Adult"
                     /lab_host="DH10B"
                     /note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;
                     Site_2: Sal 1; Starting library constructed using
                     SuperScript Plasmid Library kit (Life Technologies). cDNA
                     made by oligo-dT priming. Size-selected by column
                     fractionation; average insert size 1.08 kb. Library was
                     amplified once on solid support and plasmid DNA from
                     library was prepared. The library DNA was normalized by
                     method #4 from Bonaldo, Lennon, and Soares 1996 Genome
                     Research 6:791-806; 0.5 microgram single-stranded library
                     plasmid DNA was mixed with 5 micrograms PCR product
                     representing library inserts and hybridized to an Ecot of
                     20. Single-stranded (unhybridized) plasmids were isolated
                     by hydroxyapatite chromatography and used to make this
                     library."
BASE COUNT      185 a    140 c    118 g    138 t
ORIGIN


alignment_scores:
            Quality:  971.00            Length:    189
              Ratio:  5.138               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x BM129777   ..

 Align seg 1/1  to: BM129777  from: 1  to: 581

      18 ArgLeuThrGlnGlyArgTrpValArgLysSerArgValAlaMetGluLy 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 CGACTCACACAAGGCAGGTGGGTGAGGAAATCCAGAGTTGCCATGGAGAA 50

      34 sIleProValSerAlaPheLeuLeuLeuValAlaLeuSerTyrThrLeuA 51
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 AATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCTCCTACACTCTGG 100

      51 laArgAspThrThrValLysProGlyAlaLysLysAspThrLysAspSer 67
         |||||||||||||||||||||||||||||||||||||||||||||||||
     101 CCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGACACAAAGGACTCT 150

      68 ArgProLysLeuProGlnThrLeuSerArgGlyTrpGlyAspGlnLeuIl 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 CGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGGGTGACCAACTCAT 200

      84 eTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerLysThrSerAsnL 101
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 CTGGACTCAGACATATGAAGAAGCTCTATATAAATCCAAGACAAGCAACA 250

     101 ysProLeuMetIleIleHisHisLeuAspGluCysProHisSerGlnAla 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 AACCCTTGATGATTATTCATCACTTGGATGAGTGCCCACACAGTCAAGCT 300

     118 LeuLysLysValPheAlaGluAsnLysGluIleGlnLysLeuAlaGluGl 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     301 TTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCAGAAATTGGCAGAGCA 350

     134 nPheValLeuLeuAsnLeuValTyrGluThrThrAspLysHisLeuSerP 151
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     351 GTTTGTCCTCCTCAATCTGGTTTATGAAACAACTGACAAACACCTTTCTC 400

     151 roAspGlyGlnTyrValProArgIleMetPheValAspProSerLeuThr 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     401 CTGATGGCCAGTATGTCCCCAGGATTATGTTTGTTGACCCATCTCTGACA 450

     168 ValArgAlaAspIleThrGlyArgTyrSerAsnArgLeuTyrAlaTyrGl 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     451 GTTAGAGCCGATATCACTGGAAGATATTCAAATCGTCTCTATGCTTACGA 500

     184 uProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLysAlaLeuLysL 201
         |||||||||||||||||||||||||||||||||||||||||||||||||
     501 ACCTGCAGATACAGCTCTGTTGCTTGACAACATGAAGAAAGCTCTCAAGT 550

     201 euLeuLysThrGluLeu 206
         ||||||||||||||||||
     551 TGCTGAAGACTGAATTG 567

seq_name: gb_est2:BE796856

seq_documentation_block:
LOCUS       BE796856                948 bp    mRNA    linear   EST 20-SEP-2000
DEFINITION  601588166F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3942112 5',
            mRNA sequence.
ACCESSION   BE796856
VERSION     BE796856.1  GI:10218156
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 948)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM795  row: d  column: 17
            High quality sequence stop: 748.
FEATURES             Location/Qualifiers
     source          1. .948
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3942112"
                     /clone_lib="NIH_MGC_7"
                     /tissue_type="small cell carcinoma"
                     /cell_line="MGC3"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Size-selected >500bp for average
                     insert size 1.8kb. Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      296 a    225 c    207 g    220 t
ORIGIN


alignment_scores:
            Quality:  962.00            Length:    187
              Ratio:  5.144               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x BE796856   ..

 Align seg 1/1  to: BE796856  from: 1  to: 948

      20 ThrGlnGlyArgTrpValArgLysSerArgValAlaMetGluLysIlePr 36
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       3 ACACAAGGCAGGTGGGTGAGGAAATCCAGAGTTGCCATGGAGAAAATTCC 52
```

```
     36 oValSerAlaPheLeuLeuLeuValAlaLeuSerTyrThrLeuAlaArgA 53
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     53 AGTGTCAGCATTCTTGCTCCTTGTGGCCCTCTCCTACACTCTGGCCAGAG 102

     53 spThrThrValLysProGlyAlaLysLysAspThrLysAspSerArgPro 69
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    103 ATACCACAGTCAAACCTGGAGCCAAAAAGGACACAAAGGACTCTCGACCC 152

     70 LysLeuProGlnThrLeuSerArgGlyTrpGlyAspGlnLeuIleTrpTh 86
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    153 AAACTGCCCCAGACCCTCTCCAGAGGTTGGGGTGACCAACTCATCTGGAC 202

     86 rGlnThrTyrGluGluAlaLeuTyrLysSerLysThrSerAsnLysProL 103
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    203 TCAGACATATGAAGAAGCTCTATATAAATCCAAGACAAGCAACAAACCCT 252

    103 euMetIleIleHisHisLeuAspGluCysProHisSerGlnAlaLeuLys 119
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    253 TGATGATTATTCATCACTTGGATGAGTGCCCACACAGTCAAGCTTTAAAG 302

    120 LysValPheAlaGluAsnLysGluIleGlnLysLeuAlaGluGlnPheVa 136
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    303 AAAGTGTTTGCTGAAAATAAAGAAATCCAGAAATTGGCAGAGCAGTTTGT 352

    136 lLeuLeuAsnLeuValTyrGluThrThrAspLysHisLeuSerProAspG 153
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    353 CCTCCTCAATCTGGTTTATGAAACAACTGACAAACACCTTTCTCCTGATG 402

    153 lyGlnTyrValProArgIleMetPheValAspProSerLeuThrValArg 169
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    403 GCCAGTATGTCCCCAGGATTATGTTTGTTGACCCATCTCTGACAGTTAGA 452

    170 AlaAspIleThrGlyArgTyrSerAsnArgLeuTyrAlaTyrGluProAl 186
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    453 GCCGATATCACTGGAAGATATTCAAACCGTCTCTATGCTTACGAACCTGC 502

    186 aAspThrAlaLeuLeuLeuAspAsnMetLysLysAlaLeuLysLeuLeuL 203
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    503 AGATACAGCTCTGTTGCTTGACAACATGAAGAAAGCTCTCAAGTTGCTGA 552

    203 ysThrGluLeu 206
        ||||||||||
    553 AGACTGAATTG 563

seq_name: gb_est2:BG610804

seq_documentation_block:
LOCUS       BG610804                853 bp    mRNA    linear   EST 18-APR-2001
DEFINITION  602612442F1 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:4737755 5',
            mRNA sequence.
ACCESSION   BG610804
VERSION     BG610804.1  GI:13662175
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 853)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1606  row: d  column: 12
            High quality sequence stop: 812.

FEATURES             Location/Qualifiers
     source          1. .853
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4737755"
                     /clone_lib="NIH_MGC_60"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: prostate; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                     ); Double-stranded cDNA was prepared from cell line RNA.
                     5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).  Average
                     insert size 1.5 kb (range 0.9-4.0 kb). 14/15 colonies
                     contained inserts  by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
                     Library."
BASE COUNT      273 a    193 c    174 g    213 t
ORIGIN


alignment_scores:
            Quality:  961.00              Length:    195
              Ratio:  5.085                 Gaps:      1
 Percent Similarity:  96.923   Percent Identity: 96.923

alignment_block:
 US-09-674-266A-181 x BG610804   ..

 Align seg 1/1  to: BG610804  from: 1  to: 853

     12 GlyProHisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSe 28
        ||||||||||||||||||||||||||||||||
     14 GGACCGCATCCTAGCCGCCGACTCACACAAGGC..................  46

     28 rArgValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValA 45
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     47 .AGAGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGG 95

     45 laLeuSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLys 61
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     96 CCCTCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAA 145

     62 LysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGl 78
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    146 AAGGACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGG 195

     78 yTrpGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrL 95
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    196 TTGGGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATA 245

     95 ysSerLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGlu 111
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    246 AATCCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAG 295

    112 CysProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIl 128
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    296 TGCCCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAAT 345

    128 eGlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrT 145
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    346 CCAGAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAA 395

    145 hrAspLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPhe 161
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    396 CTGACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTT 445

    162 ValAspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAs 178
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    446 GTTGACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAA 495
```

```
    178 nArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnM 195
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    496 TCGTCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACA 545

    195 etLysLysAlaLeuLysLeuLeuLysThrGluLeu 206
        |||||||||||||||||||||||||||||||||||
    546 TGAAGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 580

seq_name: gb_est2:BE870718

seq_documentation_block:
LOCUS       BE870718                734 bp    mRNA    linear   EST 20-OCT-2000
DEFINITION  601448570F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3852448 5',
            mRNA sequence.
ACCESSION   BE870718
VERSION     BE870718.1  GI:10319507
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 734)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9574  row: l  column: 17
            High quality sequence stop: 723.
FEATURES             Location/Qualifiers
     source          1. .734
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3852448"
                     /clone_lib="NIH_MGC_65"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.8 kb.   Library constructed by Life
                     Technologies.  "
BASE COUNT      231 a    163 c    156 g    184 t
ORIGIN


alignment_scores:
            Quality:  957.00            Length:    186
              Ratio:  5.145               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x BE870718   ..

 Align seg 1/1  to: BE870718  from: 1  to: 734

      21 GlnGlyArgTrpValArgLysSerArgValAlaMetGluLysIleProVa 37
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       2 CAAGGCAGGTGGGTGAGGAAATCCAGAGTTGCCATGGAGAAAATTCCAGT 51

      37 lSerAlaPheLeuLeuLeuValAlaLeuSerTyrThrLeuAlaArgAspT 54
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      52 GTCAGCATTCTTGCTCCTTGTGGCCCTCTCCTACACTCTGGCCAGAGATA 101

      54 hrThrValLysProGlyAlaLysLysAspThrLysAspSerArgProLys 70
         ||||||||||||||||||||||||||||||||||||||||||||||||||

     102 CCACAGTCAAACCTGGAGCCAAAAAGGACACAAAGGACTCTCGACCCAAA 151

      71 LeuProGlnThrLeuSerArgGlyTrpGlyAspGlnLeuIleTrpThrGl 87
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     152 CTGCCCCAGACCCTCTCCAGAGGTTGGGGTGACCAACTCATCTGGACTCA 201

      87 nThrTyrGluGluAlaLeuTyrLysSerLysThrSerAsnLysProLeuM 104
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     202 GACATATGAAGAAGCTCTATATAAATCCAAGACAAGCAACAAACCCTTGA 251

     104 etIleIleHisHisLeuAspGluCysProHisSerGlnAlaLeuLysLys 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     252 TGATTATTCATCACTTGGATGAGTGCCCACACAGTCAAGCTTTAAAGAAA 301

     121 ValPheAlaGluAsnLysGluIleGlnLysLeuAlaGluGlnPheValLe 137
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     302 GTGTTTGCTGAAAATAAAGAAATCCAGAAATTGGCAGAGCAGTTTGTCCT 351

     137 uLeuAsnLeuValTyrGluThrThrAspLysHisLeuSerProAspGlyG 154
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     352 CCTCAATCTGGTTTATGAAACAACTGACAAACACCTTTCTCCTGATGGCC 401

     154 lnTyrValProArgIleMetPheValAspProSerLeuThrValArgAla 170
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     402 AGTATGTCCCCAGGATTATGTTTGTTGACCCATCTCTGACAGTTAGAGCC 451

     171 AspIleThrGlyArgTyrSerAsnArgLeuTyrAlaTyrGluProAlaAs 187
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     452 GATATCACTGGAAGATATTCAAACCGTCTCTATGCTTACGAACCTGCAGA 501

     187 pThrAlaLeuLeuLeuAspAsnMetLysLysAlaLeuLysLeuLeuLysT 204
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     502 TACAGCTCTGTTGCTTGACAACATGAAGAAAGCTCTCAAGTTGCTGAAGA 551

     204 hrGluLeu 206
         ||||||||
     552 CTGAATTG 559

seq_name: gb_est2:BM006356

seq_documentation_block:
LOCUS       BM006356                758 bp    mRNA    linear   EST 30-OCT-2001
DEFINITION  603615055F1 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:5421109 5',
            mRNA sequence.
ACCESSION   BM006356
VERSION     BM006356.1  GI:16520710
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 758)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1876  row: m  column: 14
            High quality sequence stop: 754.
FEATURES             Location/Qualifiers
     source          1. .758
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5421109"
                     /clone_lib="NIH_MGC_110"
                     /tissue_type="ductal carcinoma, cell line"
```

```
                    /lab_host="DH10B (phage-resistant)"
                    /note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
                    Site_2: EcoRI; cDNA made by oligo-dT priming.
                    Directionally cloned into EcoRI/XhoI sites using the
                    following 5' adaptor: GGCACGAG(G). Library constructed  by
                    Ling Hong in the laboratory of Gerald M. Rubin (University
                    of  California, Berkeley) using ZAP-cDNA synthesis kit
                    (Stratagene) and Superscript II RT (Life Technologies).
                    Note: this is a NIH_MGC Library."
BASE COUNT      238 a    176 c    156 g    188 t
ORIGIN


alignment_scores:
            Quality:  948.00             Length:    193
              Ratio:  5.070                Gaps:      1
 Percent Similarity:  96.891   Percent Identity: 96.891

alignment_block:
 US-09-674-266A-181 x BM006356   ..

 Align seg 1/1  to: BM006356  from: 1  to: 758

      14 HisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerArgVa 30
         ||||||||||||||||||||||||||                 |||||
       4 CATCCTAGCCGCCGACTCACACAAGGC.................AGAGT 35

      30 lAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAlaLeuS 47
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      36 TGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCT 85

      47 erTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLysAsp 63
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      86 CCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGAC 135

      64 ThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTrpGl 80
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     136 ACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGGG 185

      80 yAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerL 97
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     186 TGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAATCCA 235

      97 ysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCysPro 113
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     236 AGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGCCCA 285

     114 HisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGlnLy 130
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     286 CACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCAGAA 335

     130 sLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrAspL 147
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     336 ATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTGACA 385

     147 ysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheValAsp 163
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     386 AACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTTGAC 435

     164 ProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnArgLe 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     436 CCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAACCGTCT 485

     180 uTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetLysL 197
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     486 CTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACATGAAGA 535

     197 ysAlaLeuLysLeuLeuLysThrGluLeu 206
         |||||||||||||||||||||||||||||
     536 AAGCTCTCAAGTTGCTGAAGACTGAATTG 564
```

```
OM of: US-09-674-266A-181 to: GenEmbl:*   out_format : pfs

Date: Oct 8, 2002 11:42 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlh
 -Q=/cgn2_1/USPTO_spool/US09674266/runat_08102002_093409_27005/app_query.fasta_1.271
 -DB=GenEmbl -QFMT=fastap -SUFFIX=olip2n.rge -GAPOP=4.500
 -GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=quality
 -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext
 -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US09674266_@CGN1_1_3973 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-674-266A-181
Query length: 206
Database: GenEmbl:*
Database sequences: 1797656
Database length: 1873333701
Search time (sec): 1915.830000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_list:
```

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
gb_pat:AX011612	+	206.00	3944.17	2.5e-211	1020	! AX011612 Sequence 8 from Paten
gb_pat:AX015056	+	206.00	3944.17	2.5e-211	1020	! AX015056 Sequence 265 from Pat
gb_pr:AF088867	+	195.00	3732.66	1.5e-199	956	! AF088867 Homo sapiens putative
gb_pr:AF038451	+	194.00	3712.51	2.0e-198	1077	! AF038451 Homo sapiens secreted
gb_pr:AF115926	+	194.00	3709.14	3.1e-198	1701	! AF115926 Homo sapiens XAG-2 ho
gb_pr:BC015503	+	193.00	3695.10	1.9e-197	836	! BC015503 Homo sapiens, anterior
gb_pat:AR123828	+	192.00	3675.49	2.3e-196	875	! AR123828 Sequence 1 from patent
gb_pat:AX341084	+	184.00	3524.65	5.8e-188	557	! AX341084 Sequence 1331 from Pat
gb_pat:AX231601	+	178.00	3409.20	1.6e-181	543	! AX231601 Sequence 7 from Patent
gb_pr:AF007791	+	178.00	3406.00	2.4e-181	838	! AF007791 Homo sapiens secreted
gb_pat:AR177367	+	178.00	3400.83	4.6e-181	1689	! AR177367 Sequence 61 from pate
gb_pat:AX341234	+	177.00	3389.15	2.1e-180	603	! AX341234 Sequence 1481 from Pat
gb_pat:AX341953	+	161.00	3081.82	2.7e-163	525	! AX341953 Sequence 2200 from Pat
gb_pat:AX351468	-	160.00	3060.44	4.2e-162	699	! AX351468 Sequence 215 from Pate
gb_pat:AX067336	+	134.00	2558.57	3.8e-134	778	! AX067336 Sequence 40 from Paten
gb_pat:AX192616	+	133.00	2544.04	2.4e-133	409	! AX192616 Sequence 183 from Pate
gb_pat:AX341884	-	121.00	2310.67	2.4e-120	544	! AX341884 Sequence 2131 from Pat
gb_pat:AX341524	-	120.00	2291.40	2.9e-119	544	! AX341524 Sequence 1771 from Pat
gb_pat:AX341388	+	120.00	2290.76	3.1e-119	593	! AX341388 Sequence 1635 from Pat
gb_pat:AX340668	+	115.00	2195.48	6.3e-114	512	! AX340668 Sequence 915 from Pate
gb_pat:AX341186	-	108.00	2059.62	2.3e-106	583	! AX341186 Sequence 1433 from Pat
gb_pat:AR123857	+	107.00	2043.39	1.9e-105	386	! AR123857 Sequence 35 from paten
gb_pat:AX340984	-	106.00	2021.45	3.1e-104	554	! AX340984 Sequence 1231 from Pat
gb_pat:AX302898	-	103.00	1967.66	3.1e-101	321	! AX302898 Sequence 218 from Pate
gb_pat:AX156194	-	103.00	1963.98	5.0e-101	529	! AX156194 Sequence 51 from Paten
gb_pat:AR144213	+	102.00	1941.72	8.6e-100	793	! AR144213 Sequence 78 from paten
gb_pat:AR176400	+	102.00	1941.72	8.6e-100	793	! AR176400 Sequence 78 from paten
gb_pat:AX341139	+	101.00	1925.86	6.6e-99	499	! AX341139 Sequence 1386 from Pat
gb_pat:AX341299	+	93.00	1773.04	2.1e-90	415	! AX341299 Sequence 1546 from Pat
gb_pat:AX322727	+	80.00	1524.37	1.5e-76	322	! AX322727 Sequence 2 from Patent
gb_pat:AX341229	-	76.00	1443.70	4.7e-72	524	! AX341229 Sequence 1476 from Pat
gb_pat:AR166839	+	66.00	1252.95	2.0e-61	401	! AR166839 Sequence 32 from paten
gb_pat:AX192465	+	66.00	1252.95	2.0e-61	401	! AX192465 Sequence 32 from Paten
gb_pat:AR177410	+	62.00	1169.58	8.8e-57	940	! AR177410 Sequence 149 from pate
gb_pat:AX192553	+	58.00	1100.88	5.9e-53	301	! AX192553 Sequence 120 from Pate
gb_pat:AX341116	-	57.00	1076.67	1.3e-51	588	! AX341116 Sequence 1363 from Pat
gb_pat:AX261879	+	56.00	1065.32	5.6e-51	201	! AX261879 Sequence 1530 from Pat
gb_pat:AX341514	-	55.00	1042.00	1.1e-49	348	! AX341514 Sequence 1761 from Pat
gb_pat:AR123858	+	53.00	1002.94	1.7e-47	373	! AR123858 Sequence 36 from pa
gb_pat:AX106428	+	52.00	989.95	8.9e-47	159	! AX106428 Sequence 209 from P
gb_pat:AX140719	+	52.00	989.95	8.9e-47	159	! AX140719 Sequence 209 from P
gb_pat:AX200579	+	52.00	989.95	8.9e-47	159	! AX200579 Sequence 209 from P
gb_pat:AX267235	+	52.00	989.95	8.9e-47	159	! AX267235 Sequence 209 from P
gb_pat:AR123859	+	51.00	970.73	1.0e-45	158	! AR123859 Sequence 37 from pa
gb_pat:AX062783	-	50.00	948.53	1.8e-44	235	! AX062783 Sequence 410 from P

```
seq_name: gb_pat:AX011612

seq_documentation_block:
LOCUS       AX011612               1020 bp    DNA     linear   PAT 06-SEP-2000
DEFINITION  Sequence 8 from Patent WO9955858.
ACCESSION   AX011612
VERSION     AX011612.1  GI:9998136
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1020)
  AUTHORS   Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
            Pilarsky,C.
  TITLE     Human nucleic acid sequences obtained from pancreas tumor tissue
  JOURNAL   Patent: WO 9955858-A 8 04-NOV-1999;
            SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
            BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
            (DE); PILARSKY CHRISTIAN (DE)
FEATURES             Location/Qualifiers
     source          1. .1020
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      341 a    214 c    218 g    247 t
ORIGIN


alignment_scores:
            Quality:  206.00              Length:    206
              Ratio:  1.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x AX011612   ..

 Align seg 1/1  to: AX011612  from: 1  to: 1020

       1 ArgLeuSerCysAlaGlyThrLeuSerGlySerGlyProHisProSerAr 17
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      57 AGACTCAGCTGTGCTGGCACACTCAGCGGCTCTGGACCGCATCCTAGCCG 106

      17 gArgLeuThrGlnGlyArgTrpValArgLysSerArgValAlaMetGluL 34
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     107 CCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAGAGTTGCCATGGAGA 156

      34 ysIleProValSerAlaPheLeuLeuLeuValAlaLeuSerTyrThrLeu 50
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     157 AAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCTCCTACACTCTG 206

      51 AlaArgAspThrThrValLysProGlyAlaLysLysAspThrLysAspSe 67
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     207 GCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGACACAAAGGACTC 256

      67 rArgProLysLeuProGlnThrLeuSerArgGlyTrpGlyAspGlnLeuI 84
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     257 TCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGGGTGACCAACTCA 306

      84 leTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerLysThrSerAsn 100
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     307 TCTGGACTCAGACATATGAAGAAGCTCTATATAAATCCAAGACAAGCAAC 356

     101 LysProLeuMetIleIleHisHisLeuAspGluCysProHisSerGlnAl 117
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     357 AAACCCTTGATGATTATTCATCACTTGGATGAGTGCCCACACAGTCAAGC 406
```

```
     117 aLeuLysLysValPheAlaGluAsnLysGluIleGlnLysLeuAlaGluG 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     407 TTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCAGAAATTGGCAGAGC 456

     134 lnPheValLeuLeuAsnLeuValTyrGluThrThrAspLysHisLeuSer 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     457 AGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTGACAAACACCTTTCT 506

     151 ProAspGlyGlnTyrValProArgIleMetPheValAspProSerLeuTh 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     507 CCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTTGACCCATCTCTGAC 556

     167 rValArgAlaAspIleThrGlyArgTyrSerAsnArgLeuTyrAlaTyrG 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     557 AGTTAGAGCCGATATCACTGGAAGATATTCAAACCGTCTCTATGCTTACG 606

     184 luProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLysAlaLeuLys 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     607 AACCTGCAGATACAGCTCTGTTGCTTGACAACATGAAGAAAGCTCTCAAG 656

     201 LeuLeuLysThrGluLeu 206
         |||||||||||||||||
     657 TTGCTGAAGACTGAATTG 674

seq_name: gb_pat:AX015056

seq_documentation_block:
LOCUS       AX015056               1020 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 265 from Patent WO9953040.
ACCESSION   AX015056
VERSION     AX015056.1  GI:10041195
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1020)
  AUTHORS   Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
            Pilarsky,C.
  TITLE     Human nucleic acid sequences from ovarian tumour tissue
  JOURNAL   Patent: WO 9953040-A 265 21-OCT-1999;
            SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
            BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
            (DE); PILARSKY CHRISTIAN (DE)
FEATURES             Location/Qualifiers
     source          1. .1020
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      341 a    214 c    218 g    247 t
ORIGIN


alignment_scores:
            Quality:  206.00             Length:    206
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x AX015056   ..

 Align seg 1/1  to: AX015056  from: 1  to: 1020

       1 ArgLeuSerCysAlaGlyThrLeuSerGlySerGlyProHisProSerAr 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      57 AGACTCAGCTGTGCTGGCACACTCAGCGGCTCTGGACCGCATCCTAGCCG 106

      17 gArgLeuThrGlnGlyArgTrpValArgLysSerArgValAlaMetGluL 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     107 CCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAGAGTTGCCATGGAGA 156

      34 ysIleProValSerAlaPheLeuLeuLeuValAlaLeuSerTyrThrLeu 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     157 AAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCTCCTACACTCTG 206

      51 AlaArgAspThrThrValLysProGlyAlaLysLysAspThrLysAspSe 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     207 GCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGACACAAAGGACTC 256

      67 rArgProLysLeuProGlnThrLeuSerArgGlyTrpGlyAspGlnLeuI 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     257 TCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGGGTGACCAACTCA 306

      84 leTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerLysThrSerAsn 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     307 TCTGGACTCAGACATATGAAGAAGCTCTATATAAATCCAAGACAAGCAAC 356

     101 LysProLeuMetIleIleHisHisLeuAspGluCysProHisSerGlnAl 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     357 AAACCCTTGATGATTATTCATCACTTGGATGAGTGCCCACACAGTCAAGC 406

     117 aLeuLysLysValPheAlaGluAsnLysGluIleGlnLysLeuAlaGluG 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     407 TTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCAGAAATTGGCAGAGC 456

     134 lnPheValLeuLeuAsnLeuValTyrGluThrThrAspLysHisLeuSer 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     457 AGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTGACAAACACCTTTCT 506

     151 ProAspGlyGlnTyrValProArgIleMetPheValAspProSerLeuTh 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     507 CCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTTGACCCATCTCTGAC 556

     167 rValArgAlaAspIleThrGlyArgTyrSerAsnArgLeuTyrAlaTyrG 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     557 AGTTAGAGCCGATATCACTGGAAGATATTCAAACCGTCTCTATGCTTACG 606

     184 luProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLysAlaLeuLys 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     607 AACCTGCAGATACAGCTCTGTTGCTTGACAACATGAAGAAAGCTCTCAAG 656

     201 LeuLeuLysThrGluLeu 206
         |||||||||||||||||
     657 TTGCTGAAGACTGAATTG 674

seq_name: gb_pr:AF088867

seq_documentation_block:
LOCUS       AF088867               956 bp    mRNA    linear   PRI 02-JAN-2000
DEFINITION  Homo sapiens putative secreted protein XAG mRNA, complete cds.
ACCESSION   AF088867
VERSION     AF088867.1  GI:6652811
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 956)
  AUTHORS   Zhang,J.S. and Smith,D.I.
  TITLE     Human homolog of XAG is differentially expressed in tumors
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 956)
  AUTHORS   Zhang,J.S. and Smith,D.I.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-AUG-1998) Pathology and Lab. Medicine, Mayo Clinic,
            200 SW 1st St., Rochester, MN 55905, USA
FEATURES             Location/Qualifiers
     source          1. .956
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             148. .675
                     /note="differentially expressed in tumors; similar to
                     Xenopus cement gland-specific secreted protein"
                     /codon_start=1
                     /product="putative secreted protein XAG"
                     /protein_id="AAF22484.1"
```

```
                     /db_xref="GI:6652812"
                     /translation="MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTL
                     SRGWGDQLIWTQTYEEALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAE
                     QFVLLNLVYETTDKHLSPDGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALL
                     LDNMKKALKLLKTEL"
BASE COUNT      310 a    202 c    202 g    242 t
ORIGIN


alignment_scores:
            Quality:  195.00             Length:    195
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x AF088867   ..

 Align seg 1/1  to: AF088867  from: 1  to: 956

      12 GlyProHisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSe 28
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      88 GGACCGCATCCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATC 137

      28 rArgValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValA 45
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     138 CAGAGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGG 187

      45 laLeuSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLys 61
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     188 CCCTCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAA 237

      62 LysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGl 78
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     238 AAGGACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGG 287

      78 yTrpGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrL 95
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     288 TTGGGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATA 337

      95 ysSerLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGlu 111
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     338 AATCCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAG 387

     112 CysProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIl 128
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     388 TGCCCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAAT 437

     128 eGlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrT 145
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     438 CCAGAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAA 487

     145 hrAspLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPhe 161
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     488 CTGACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTT 537

     162 ValAspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAs 178
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     538 GTTGACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAA 587

     178 nArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnM 195
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     588 TCGTCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACA 637

     195 etLysLysAlaLeuLysLeuLeuLysThrGluLeu 206
         |||||||||||||||||||||||||||||||||||
     638 TGAAGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 672

seq_name: gb_pr:AF038451

seq_documentation_block:
LOCUS       AF038451               1077 bp    mRNA    linear   PRI 03-DEC-1998
DEFINITION  Homo sapiens secreted cement gland protein XAG-2 homolog (hAG-2/R)
            mRNA, complete cds.
ACCESSION   AF038451
VERSION     AF038451.1  GI:3779225
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1077)
  AUTHORS   Thompson,D.A. and Weigel,R.J.
  TITLE     hAG-2, the human homologue of the Xenopus laevis cement gland gene
            XAG-2, is coexpressed with estrogen receptor in breast cancer cell
            lines
  JOURNAL   Biochem. Biophys. Res. Commun. 251 (1), 111-116 (1998)
  MEDLINE   99009231
REFERENCE   2  (bases 1 to 1077)
  AUTHORS   Thompson,D.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1997) Surgery, Stanford University, 1201 Welch
            Road, MSLS Building, Room P228, Stanford, CA 94305, USA
FEATURES             Location/Qualifiers
     source          1. .1077
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="MCF7"
                     /tissue_type="breast"
                     /note="isolated from estrogen receptor-positive breast
                     adenocarcinoma cell line"
     gene            1. .1077
                     /gene="hAG-2/R"
     CDS             59. .586
                     /gene="hAG-2/R"
                     /note="similar to Homo sapiens protein hAG-2/C deposited
                     under GenBank Accession Number AF007791, and to Xenopus
                     laevis cement gland-specific proteins np77 and XAG-2
                     deposited under GenBank Accession Numbers U82110 and
                     AF025474, respectively"
                     /codon_start=1
                     /product="secreted cement gland protein XAG-2 homolog"
                     /protein_id="AAC82614.1"
                     /db_xref="GI:3779226"
                     /translation="MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTL
                     SRGWGDQLIWTQTYEEALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAE
                     QFVLLNLVYETTDKHLSPDGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALL
                     LDNMKKALKLLKTEL"
BASE COUNT      367 a    231 c    210 g    269 t
ORIGIN


alignment_scores:
            Quality:  194.00             Length:    194
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x AF038451   ..

 Align seg 1/1  to: AF038451  from: 1  to: 1077

      13 ProHisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerAr 29
         |||||||||||||||||||||||||||||||||||||||||||||||||||
       2 CCGCATCCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAG 51

      29 gValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAlaL 46
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      52 AGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCC 101

      46 euSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLys 62
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     102 TCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAG 151

      63 AspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTr 79
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     152 GACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTG 201
```

```
     79 pGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysS 96
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    202 GGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAAT 251

     96 erLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCys 112
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    252 CCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGC 301

    113 ProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGl 129
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    302 CCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCA 351

    129 nLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrA 146
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    352 GAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTG 401

    146 spLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheVal 162
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    402 ACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTT 451

    163 AspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnAr 179
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    452 GACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAATCG 501

    179 gLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetL 196
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    502 TCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACATGA 551

    196 ysLysAlaLeuLysLeuLeuLysThrGluLeu 206
        |||||||||||||||||||||||||||||||
    552 AGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 583

seq_name: gb_pr:AF115926

seq_documentation_block:
LOCUS       AF115926               1701 bp    mRNA    linear   PRI 30-DEC-2001
DEFINITION  Homo sapiens XAG-2 homolog long protein (HPC8) mRNA, complete cds.
ACCESSION   AF115926
VERSION     AF115926.1  GI:17998664
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1701)
  AUTHORS   Zhang,J.S. and Smith,D.I.
  TITLE     Identification of human homolog of XAG-2 over-expressed in tumors
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1701)
  AUTHORS   Zhang,J.S. and Smith,D.I.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-DEC-1998) Laboratory of Medicine and Pathology, Mayo
            Clinic, 200 SW 1st Street, Rochester, MN 55905, USA
FEATURES             Location/Qualifiers
     source          1. .1701
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_type="prostate cancer"
     gene            1. .1701
                     /gene="HPC8"
     CDS             58. .585
                     /gene="HPC8"
                     /codon_start=1
                     /product="XAG-2 homolog long protein"
                     /protein_id="AAL54870.1"
                     /db_xref="GI:17998665"
                     /translation="MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTL
                     SRGWGDQLIWTQTYEEALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAE
                     QFVLLNLVYETTDKHLSPDGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALL
                     LDNMKKALKLLKTEL"
BASE COUNT      557 a    352 c    339 g    453 t
ORIGIN
```

```
alignment_scores:
            Quality:  194.00              Length:    194
              Ratio:  1.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x AF115926   ..

 Align seg 1/1  to: AF115926  from: 1  to: 1701

     13 ProHisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerAr 29
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 CCGCATCCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAG 50

     29 gValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAlaL 46
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     51 AGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCC 100

     46 euSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLys 62
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    101 TCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAG 150

     63 AspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTr 79
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    151 GACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTG 200

     79 pGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysS 96
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    201 GGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAAT 250

     96 erLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCys 112
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    251 CCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGC 300

    113 ProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGl 129
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    301 CCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCA 350

    129 nLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrA 146
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    351 GAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTG 400

    146 spLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheVal 162
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    401 ACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTT 450

    163 AspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnAr 179
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    451 GACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAATCG 500

    179 gLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetL 196
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    501 TCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACATGA 550

    196 ysLysAlaLeuLysLeuLeuLysThrGluLeu 206
        |||||||||||||||||||||||||||||||
    551 AGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 582

seq_name: gb_pr:BC015503

seq_documentation_block:
LOCUS       BC015503               836 bp    mRNA    linear   PRI 29-OCT-2001
DEFINITION  Homo sapiens, anterior gradient 2 (Xenepus laevis) homolog, clone
            MGC:9112 IMAGE:3852448, mRNA, complete cds.
ACCESSION   BC015503
VERSION     BC015503.1  GI:15930124
KEYWORDS    MGC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE   1  (bases 1 to 836)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:       http://www-shgc.stanford.edu
            Contact:  (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 20 Row: o Column: 11
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 6652811.
FEATURES             Location/Qualifiers
     source          1. .836
                     /organism="Homo sapiens"
                     /db_xref="LocusID:10551"
                     /db_xref="taxon:9606"
                     /clone="MGC:9112 IMAGE:3852448"
                     /tissue_type="Colon, adenocarcinoma"
                     /clone_lib="NIH_MGC_65"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS             57. .584
                     /codon_start=1
                     /product="anterior gradient 2 (Xenepus laevis) homolog"
                     /protein_id="AAH15503.1"
                     /db_xref="GI:15930125"
                     /translation="MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTL
                     SRGWGDQLIWTQTYEEALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAE
                     QFVLLNLVYETTDKHLSPDGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALL
                     LDNMKKALKLLKTEL"
BASE COUNT      282 a    181 c    166 g    207 t
ORIGIN


alignment_scores:
            Quality:  193.00            Length:    193
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x BC015503   ..

 Align seg 1/1  to: BC015503  from: 1  to: 836

      14 HisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerArgVa 30
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       3 CATCCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAGAGT 52

      30 lAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAlaLeuS 47
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      53 TGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCT 102

      47 erTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLysAsp 63
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     103 CCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGAC 152

      64 ThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTrpGl 80
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     153 ACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGGG 202

      80 yAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerL 97
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     203 TGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAATCCA 252

      97 ysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCysPro 113
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     253 AGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGCCCA 302

     114 HisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGlnLy 130
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     303 CACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCAGAA 352

     130 sLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrAspL 147
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     353 ATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTGACA 402

     147 ysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheValAsp 163
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     403 AACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTTGAC 452

     164 ProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnArgLe 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     453 CCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAACCGTCT 502

     180 uTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetLysL 197
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     503 CTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACATGAAGA 552

     197 ysAlaLeuLysLeuLeuLysThrGluLeu 206
         |||||||||||||||||||||||||||||
     553 AAGCTCTCAAGTTGCTGAAGACTGAATTG 581

seq_name: gb_pat:AR123828

seq_documentation_block:
LOCUS       AR123828                875 bp    DNA     linear   PAT 16-MAY-2001
DEFINITION  Sequence 1 from patent US 6171816.
ACCESSION   AR123828
VERSION     AR123828.1  GI:14109189
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 875)
  AUTHORS   Yu,G., Dillon,P.J., Ebner,R. and Endress,G.A.
  TITLE     Human XAG-1 polynucleotides and polypeptides
  JOURNAL   Patent: US 6171816-A 1 09-JAN-2001;
FEATURES             Location/Qualifiers
     source          1. .875
                     /organism="unknown"
BASE COUNT      283 a    189 c    180 g    223 t
ORIGIN


alignment_scores:
            Quality:  192.00            Length:    192
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x AR123828   ..

 Align seg 1/1  to: AR123828  from: 1  to: 875

      15 ProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerArgValAl 31
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      20 CCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAGAGTTGC 69

      31 aMetGluLysIleProValSerAlaPheLeuLeuLeuValAlaLeuSerT 48
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      70 CATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCTCCT 119
```

```
     48 yrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLysAspThr 64
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    120 ACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGACACA 169

     65 LysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTrpGlyAs 81
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    170 AAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGGGTGA 219

     81 pGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerLysT 98
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    220 CCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAATCCAAGA 269

     98 hrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCysProHis 114
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    270 CAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGCCCACAC 319

    115 SerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGlnLysLe 131
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    320 AGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCAGAAATT 369

    131 uAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrAspLysH 148
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    370 GGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTGACAAAC 419

    148 isLeuSerProAspGlyGlnTyrValProArgIleMetPheValAspPro 164
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    420 ACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTTGACCCA 469

    165 SerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnArgLeuTy 181
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    470 TCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAATCGTCTCTA 519

    181 rAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLysA 198
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    520 TGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACATGAAGAAAG 569

    198 laLeuLysLeuLeuLysThrGluLeu 206
        |||||||||||||||||||||||||
    570 CTCTCAAGTTGCTGAAGACTGAATTG 595

seq_name: gb_pat:AX341084

seq_documentation_block:
LOCUS       AX341084                557 bp    DNA     linear   PAT 10-JAN-2002
DEFINITION  Sequence 1331 from Patent WO0196388.
ACCESSION   AX341084
VERSION     AX341084.1  GI:18137066
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Jiang,Y., Harlocker,S.L. and Secrist,H.
  TITLE     Compositions and methods for the therapy and diagnosis of colon
            cancer
  JOURNAL   Patent: WO 0196388-A 1331 20-DEC-2001;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .557
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      169 a    143 c    113 g    130 t      2 others
ORIGIN


alignment_scores:
            Quality:  184.00             Length:    184
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x AX341084    ..
```

```
Align seg 1/1  to: AX341084  from: 1  to: 557

     13 ProHisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerAr 29
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 CCGCATCCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAG 50

     29 gValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAlaL 46
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     51 AGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCC 100

     46 euSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLys 62
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    101 TCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAG 150

     63 AspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTr 79
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    151 GACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTG 200

     79 pGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysS 96
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    201 GGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAAT 250

     96 erLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCys 112
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    251 CCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGC 300

    113 ProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGl 129
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    301 CCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCA 350

    129 nLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrA 146
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    351 GAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTG 400

    146 spLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheVal 162
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    401 ACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTT 450

    163 AspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnAr 179
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    451 GACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAATCG 500

    179 gLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetL 196
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    501 TCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACATGA 550

    196 ys 196
        ||
    551 AG 552

seq_name: gb_pat:AX231601

seq_documentation_block:
LOCUS       AX231601                543 bp    DNA     linear   PAT 11-SEP-2001
DEFINITION  Sequence 7 from Patent WO0163290.
ACCESSION   AX231601
VERSION     AX231601.1  GI:15592500
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 543)
  AUTHORS   Boyd,R.S., Stamps,A.C., Terrett,J.A. and Tyson,K.L.
  TITLE     Bcmp-7 as marker for diagnosis of breast cancer
  JOURNAL   Patent: WO 0163290-A 7 30-AUG-2001;
            Oxford GlycoSciences (UK) Limited (GB)
FEATURES             Location/Qualifiers
     source          1. .543
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             11. .538
```

```
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAC69703.1"
                     /db_xref="GI:15592501"
                     /translation="MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTL
                     SRGWGDQLIWTQTYEEALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAE
                     QFVLLNLVYETTDKHLSPDGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALL
                     LDNMKKALKLLKTEL"
BASE COUNT      170 a    135 c    107 g    131 t
ORIGIN


alignment_scores:
            Quality:  178.00             Length:    178
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x AX231601   ..

 Align seg 1/1  to: AX231601  from: 1  to: 543

      29 ArgValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAl 45
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       2 AGAGTTGCCATGGAGAAAATCCCAGTGTCAGCATTCTTGCTCCTTGTGGC 51

      45 aLeuSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysL 62
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      52 CCTCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAA 101

      62 ysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGly 78
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     102 AGGACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGT 151

      79 TrpGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLy 95
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     152 TGGGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAA 201

      95 sSerLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluC 112
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     202 ATCCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGT 251

     112 ysProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIle 128
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     252 GCCCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATC 301

     129 GlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrTh 145
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     302 CAGAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAAC 351

     145 rAspLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheV 162
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     352 TGACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTG 401

     162 alAspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsn 178
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     402 TTGACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAAC 451

     179 ArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMe 195
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     452 CGTCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACAT 501

     195 tLysLysAlaLeuLysLeuLeuLysThrGluLeu 206
         |||||||||||||||||||||||||||||||||
     502 GAAGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 535

seq_name: gb_pr:AF007791

seq_documentation_block:
LOCUS       AF007791                838 bp    mRNA    linear   PRI 14-NOV-1998
DEFINITION  Homo sapiens secreted cement gland protein XAG-2 homolog (hAG-2/C)
            mRNA, complete cds.
ACCESSION   AF007791
VERSION     AF007791.1  GI:3779196
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 838)
  AUTHORS   Thompson,D.A. and Weigel,R.J.
  TITLE     hAG-2, the human homologue of the Xenopus laevis cement gland gene
            XAG-2, is coexpressed with estrogen receptor in breast cancer cell
            lines
  JOURNAL   Biochem. Biophys. Res. Commun. 251 (1), 111-116 (1998)
  MEDLINE   99009231
REFERENCE   2  (bases 1 to 838)
  AUTHORS   Thompson,D.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUN-1997) Surgery, Stanford, 1201 Welch Road, MSLS
            Building, Room P228, Stanford, CA 94305, USA
FEATURES             Location/Qualifiers
     source          1. .838
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="MCF7; estrogen receptor-expressing"
                     /cell_type="adenocarcinoma"
                     /tissue_type="breast"
     gene            1. .838
                     /gene="hAG-2/C"
     CDS             11. .538
                     /gene="hAG-2/C"
                     /note="similar to Homo sapiens protein hAG-2/R deposited
                     in GenBank Accession Number AF038451, and to Xenopus
                     laevis cement gland proteins np77 deposited in GenBank
                     Accession Number U82110 and XAG-2 deposited in GenBank
                     Accession Number AF025474"
                     /codon_start=1
                     /product="secreted cement gland protein XAG-2 homolog"
                     /protein_id="AAC77358.1"
                     /db_xref="GI:3779197"
                     /translation="MEKIPVSAFLLLVALSYTLARDTTVKPGAKKDTKDSRPKLPQTL
                     SRGWGDQLIWTQTYEEALYKSKTSNKPLMIIHHLDECPHSQALKKVFAENKEIQKLAE
                     QFVLLNLVYETTDKHLSPDGQYVPRIMFVDPSLTVRADITGRYSNRLYAYEPADTALL
                     LDNMKKALKLLKTEL"
BASE COUNT      291 a    170 c    160 g    217 t
ORIGIN


alignment_scores:
            Quality:  178.00             Length:    178
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x AF007791   ..

 Align seg 1/1  to: AF007791  from: 1  to: 838

      29 ArgValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAl 45
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       2 AGAGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGC 51

      45 aLeuSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysL 62
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      52 CCTCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAA 101

      62 ysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGly 78
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     102 AGGACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGT 151

      79 TrpGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLy 95
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     152 TGGGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAA 201

      95 sSerLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluC 112
```

```
     202 ATCCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGT 251

     112 ysProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIle 128
     252 GCCCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATC 301

     129 GlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrTh 145
     302 CAGAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAAC 351

     145 rAspLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheV 162
     352 TGACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTG 401

     162 alAspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsn 178
     402 TTGACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAAT 451

     179 ArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMe 195
     452 CGTCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACAT 501

     195 tLysLysAlaLeuLysLeuLeuLysThrGluLeu 206
     502 GAAGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 535

seq_name: gb_pat:AR177367

seq_documentation_block:
LOCUS       AR177367               1689 bp    DNA     linear   PAT 17-DEC-2001
DEFINITION  Sequence 61 from patent US 6312922.
ACCESSION   AR177367
VERSION     AR177367.1  GI:17919722
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1689)
  AUTHORS   Edwards,J.-B.Dumas.Milne., Duclert,A. and Bougueleret,L.
  TITLE     Complementary DNAs
  JOURNAL   Patent: US 6312922-A 61 06-NOV-2001;
FEATURES             Location/Qualifiers
     source          1. .1689
                     /organism="unknown"
BASE COUNT      552 a    350 c    335 g    452 t
ORIGIN


alignment_scores:
            Quality:  178.00             Length:    178
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x AR177367   ..

 Align seg 1/1  to: AR177367  from: 1  to: 1689

      29 ArgValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAl 45
      42 AGAGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGC 91

      45 aLeuSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysL 62
      92 CCTCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAA 141

      62 ysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGly 78
     142 AGGACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGT 191

      79 TrpGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLy 95
     192 TGGGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAA 241

      95 sSerLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluC 112
     242 ATCCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGT 291

     112 ysProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIle 128
     292 GCCCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATC 341

     129 GlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrTh 145
     342 CAGAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAAC 391

     145 rAspLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheV 162
     392 TGACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTG 441

     162 alAspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsn 178
     442 TTGACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAAT 491

     179 ArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMe 195
     492 CGTCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACAT 541

     195 tLysLysAlaLeuLysLeuLeuLysThrGluLeu 206
     542 GAAGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 575

seq_name: gb_pat:AX341234

seq_documentation_block:
LOCUS       AX341234               603 bp    DNA     linear   PAT 10-JAN-2002
DEFINITION  Sequence 1481 from Patent WO0196388.
ACCESSION   AX341234
VERSION     AX341234.1  GI:18137216
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Jiang,Y., Harlocker,S.L. and Secrist,H.
  TITLE     Compositions and methods for the therapy and diagnosis of colon
            cancer
  JOURNAL   Patent: WO 0196388-A 1481 20-DEC-2001;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .603
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      187 a    153 c    116 g    144 t      3 others
ORIGIN


alignment_scores:
            Quality:  177.00             Length:    177
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x AX341234   ..

 Align seg 1/1  to: AX341234  from: 1  to: 603

      29 ArgValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAl 45
      30 AGAGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGC 79

      45 aLeuSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysL 62
      80 CCTCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAA 129
```

```
     62 ysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGly 78
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    130 AGGACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGT 179

     79 TrpGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLy 95
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    180 TGGGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAA 229

     95 sSerLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluC 112
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    230 ATCCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGT 279

    112 ysProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIle 128
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    280 GCCCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATC 329

    129 GlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrTh 145
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    330 CAGAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAAC 379

    145 rAspLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheV 162
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    380 TGACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTG 429

    162 alAspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsn 178
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    430 TTGACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAAT 479

    179 ArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMe 195
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    480 CGTCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACAT 529

    195 tLysLysAlaLeuLysLeuLeuLysThrGlu 205
        |||||||||||||||||||||||||||||||
    530 GAAGAAAGCTCTCAAGTTGCTGAAGACTGAA 560

seq_name: gb_pat:AX341953

seq_documentation_block:
LOCUS       AX341953                525 bp    DNA     linear   PAT 10-JAN-2002
DEFINITION  Sequence 2200 from Patent WO0196388.
ACCESSION   AX341953
VERSION     AX341953.1  GI:18137935
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Jiang,Y., Harlocker,S.L. and Secrist,H.
  TITLE     Compositions and methods for the therapy and diagnosis of colon
            cancer
  JOURNAL   Patent: WO 0196388-A 2200 20-DEC-2001;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .525
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      156 a    140 c    102 g    125 t      2 others
ORIGIN


alignment_scores:
            Quality:  161.00            Length:    161
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x AX341953   ..

 Align seg 1/1  to: AX341953  from: 1  to: 525
```

```
     29 ArgValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAl 45
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     31 AGAGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGC 80

     45 aLeuSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysL 62
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     81 CCTCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAA 130

     62 ysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGly 78
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    131 AGGACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGT 180

     79 TrpGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLy 95
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    181 TGGGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAA 230

     95 sSerLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluC 112
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    231 ATCCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGT 280

    112 ysProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIle 128
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    281 GCCCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATC 330

    129 GlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrTh 145
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    331 CAGAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAAC 380

    145 rAspLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheV 162
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    381 TGACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTG 430

    162 alAspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsn 178
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    431 TTGACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAAC 480

    179 ArgLeuTyrAlaTyrGluProAlaAspThrAla 189
        |||||||||||||||||||||||||||||||||
    481 CGTCTCTATGCTTACGAACCTGCAGATACAGCT 513

seq_name: gb_pat:AX351468

seq_documentation_block:
LOCUS       AX351468                699 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 215 from Patent WO0196390.
ACCESSION   AX351468
VERSION     AX351468.1  GI:18616815
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Jiang,Y., Hepler,W.T., Clapper,J.D., Wang,A. and Secrist,H.
  TITLE     Compositions and methods for the therapy and diagnosis of colon
            cancer
  JOURNAL   Patent: WO 0196390-A 215 20-DEC-2001;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .699
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      181 a    139 c    154 g    224 t      1 others
ORIGIN


alignment_scores:
            Quality:  160.00            Length:    160
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x AX351468/rev   ..
```

```
Align seg 1/1  to reverse of: AX351468  from: 1  to: 699

     47 SerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLysAs 63
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    658 TCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGA 609

     63 pThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTrpG 80
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    608 CACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGG 559

     80 lyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysSer 96
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    558 GTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAATCC 509

     97 LysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCysPr 113
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    508 AAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGCCC 459

    113 oHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGlnL 130
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    458 ACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCAGA 409

    130 ysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrAsp 146
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    408 AATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTGAC 359

    147 LysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheValAs 163
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    358 AAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTTGA 309

    163 pProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnArgL 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    308 CCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAATCGTC 259

    180 euTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetLys 196
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    258 TCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACATGAAG 209

    197 LysAlaLeuLysLeuLeuLysThrGluLeu 206
        |||||||||||||||||||||||||||||||
    208 AAAGCTCTCAAGTTGCTGAAGACTGAATTG 179

seq_name: gb_pat:AX067336

seq_documentation_block:
LOCUS       AX067336                778 bp    DNA     linear   PAT 24-JAN-2001
DEFINITION  Sequence 40 from Patent WO0078960.
ACCESSION   AX067336
VERSION     AX067336.1  GI:12544960
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 778)
  AUTHORS   Yuqiu,J. and Mitcham,J.L.
  TITLE     Compositions and methods for the therapy and diagnosis of breast
            cancer
  JOURNAL   Patent: WO 0078960-A 40 28-DEC-2000;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .778
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     misc_feature    1. .778
                     /note="n = A,T,C or G"
BASE COUNT      245 a    175 c    158 g    193 t      7 others
ORIGIN


alignment_scores:
            Quality:  134.00            Length:    134
                Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x AX067336   ..

 Align seg 1/1  to: AX067336  from: 1  to: 778

     54 ThrThrValLysProGlyAlaLysLysAspThrLysAspSerArgProLy 70
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    125 ACCACAGTCAAACCTGGAGCCAAAAAGGACACAAAGGACTCTCGACCCAA 174

     70 sLeuProGlnThrLeuSerArgGlyTrpGlyAspGlnLeuIleTrpThrG 87
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    175 ACTGCCCCAGACCCTCTCCAGAGGTTGGGGTGACCAACTCATCTGGACTC 224

     87 lnThrTyrGluGluAlaLeuTyrLysSerLysThrSerAsnLysProLeu 103
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    225 AGACATATGAAGAAGCTCTATATAAATCCAAGACAAGCAACAAACCCTTG 274

    104 MetIleIleHisHisLeuAspGluCysProHisSerGlnAlaLeuLysLy 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    275 ATGATTATTCATCACTTGGATGAGTGCCCACACAGTCAAGCTTTAAAGAA 324

    120 sValPheAlaGluAsnLysGluIleGlnLysLeuAlaGluGlnPheValL 137
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    325 AGTGTTTGCTGAAAATAAAGAAATCCAGAAATTGGCAGAGCAGTTTGTCC 374

    137 euLeuAsnLeuValTyrGluThrThrAspLysHisLeuSerProAspGly 153
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    375 TCCTCAATCTGGTTTATGAAACAACTGACAAACACCTTTCTCCTGATGGC 424

    154 GlnTyrValProArgIleMetPheValAspProSerLeuThrValArgAl 170
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    425 CAGTATGTCCCCAGGATTATGTTTGTTGACCCATCTCTGACAGTTAGAGC 474

    170 aAspIleThrGlyArgTyrSerAsnArgLeuTyrAlaTyrGluProAlaA 187
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    475 CGATATCACTGGAAGATATTCAAATCGTCTCTATGCTTACGAACCTGCAG 524

    187 sp 187
        ||
    525 AT 526
```

OM of: US-09-674-266A-181 to: N_Geneseq_032802:* out_format : pfs

Date: Oct 8, 2002 11:46 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09674266/runat_08102002_093410_27035/app_query.fasta_1.271
-DB=N_Geneseq_032802 -QFMT=fastap -SUFFIX=olip2n.rng
-GAPOP=4.500 -GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000
-LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000
-XGAPEXT=60.000 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000
-YGAPEXT=60.000 -DELOP=6.000 -DELEXT=7.000 -START=1
-MATRIX=oligo -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
-THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09674266_@CGN1_1_0 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-674-266A-181
Query length: 206
Database: N_Geneseq_032802:*
Database sequences: 1736436
Database length: 858457221
Search time (sec): 211.730000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..			
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAZ52865	+	206.00	3721.06	6.5e-199	10				
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:AAV29047	+	195.00	3522.39	7.5e-188	86				
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAC58380	+	195.00	3521.94	8.0e-188	92				
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH35026	+	194.00	3503.59	8.4e-187	9				
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAZ77486	+	194.00	3503.01	9.0e-187	10				
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:AAV19155	+	192.00	3467.80	8.2e-185	87				
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAF63314	+	192.00	3467.80	8.2e-185	8				
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:AAV59320	+	191.00	3449.58	8.5e-184	88				
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAS13480	+	178.00	3216.97	7.7e-171	5				
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAZ40803	+	178.00	3208.49	2.3e-170	16				
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:AAV29048	+	175.00	3162.71	8.1e-168	52				
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAC00115	+	146.00	2636.59	1.6e-138	46				
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAF44884	+	134.00	2414.74	3.7e-126	7				
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAA77903	+	133.00	2401.37	2.1e-125	40				
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAI28641	+	133.00	2401.37	2.1e-125	4				
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAX41145	+	119.00	2147.18	3.0e-111	39				
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAZ42286	+	118.00	2127.01	4.0e-110	52				
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAL19619	+	112.00	2020.60	3.4e-104	3				
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:AAV19188	+	107.00	1929.35	4.1e-99	38				
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAF63347	+	107.00	1929.35	4.1e-99	3				
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAS47160	-	103.00	1858.04	3.8e-95	3				
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAF17730	-	103.00	1858.04	3.8e-95	3				
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH55526	-	103.00	1854.31	6.1e-95	5				
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAZ24578	+	102.00	1833.12	9.3e-94	79				
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAC65817	+	102.00	1833.12	9.3e-94	79				
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:ABK09584	-	101.00	1822.03	3.9e-93	30				
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAL20462	+	98.00	1760.39	1.0e-89	7				
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAL23173	-	85.00	1522.96	1.8e-76	9				
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:ABK15919	+	84.00	1512.88	6.4e-76	31				
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAL10039	+	76.00	1365.25	1.1e-67	4				
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAL11564	+	76.00	1361.48	1.7e-67	7				
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAA77753	+	66.00	1184.04	1.3e-57	40				
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAI28491	+	66.00	1184.04	1.3e-57	4				
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAZ40846	+	62.00	1104.99	3.4e-53	94				
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAC65978	+	60.00	1074.96	1.6e-51	40				
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAA77841	+	58.00	1040.81	1.3e-49	30				
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAI28579	+	58.00	1040.81	1.3e-49	3				
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAX40603	+	58.00	1039.81	1.4e-49	34				
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS58854	+	56.00	1007.49	9.1e-48					
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:AAV19189	+	53.00	948.35	1.8e-44					
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAF63348	+	53.00	948.35	1.8e-44					
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:AAV61263	+	52.00	936.55	8.1e-44					
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:AAV58679	+	52.00	936.55	8.1e-44					
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAA06442	+	52.00	936.55	8.1e-44					
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAS63650	+	52.00	936.55	8.1e-44					

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAZ52865

seq_documentation_block:
ID AAZ52865 standard; cDNA; 1020 BP.
XX
AC AAZ52865;
XX
DT 14-MAR-2000 (first entry)
XX
DE Human prostate tumor cDNA library derived EST fragment #8.
XX
KW Pancreas; tumor; EST; expressed sequence tag; human; cytostatic;
KW treatment; ds.
XX
OS Homo sapiens.
XX
PN DE19820190-A1.
XX
PD 04-NOV-1999.
XX
PF 28-APR-1998; 98DE-1020190.
XX
PR 28-APR-1998; 98DE-1020190.
XX
PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
XX
DR WPI; 1999-621386/54.
DR P-PSDB; AAY73835, AAY73836, AAY73837.
XX
PT New human nucleic acid sequences from pancreatic tumors, and related
PT proteins -
XX
PS Claim 2; Page 188; 502pp; German.
XX
CC This invention describes novel polypeptides and their encoding nucleic
CC acids derived from human pancreatic tumor tissue which have cytostatic
CC activity. The sequences are also useful in producing pharmaceutical
CC compositions for treatment of pancreatic tumors. AAZ52858-Z53014
CC represent expressed sequence tag (EST) fragments derived from a human
CC pancreatic tumor cDNA library and which encode the proteins represented
CC in AAY73814-Y74252.
XX
SQ Sequence 1020 BP; 341 A; 214 C; 218 G; 247 T; 0 other;

alignment_scores:
Quality: 206.00 Length: 206
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-266A-181 x AAZ52865 ..

Align seg 1/1 to: AAZ52865 from: 1 to: 1020

```
      1 ArgLeuSerCysAlaGlyThrLeuSerGlySerGlyProHisProSerAr 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     57 AGACTCAGCTGTGCTGGCACACTCAGCGGCTCTGGACCGCATCCTAGCCG 106

     17 gArgLeuThrGlnGlyArgTrpValArgLysSerArgValAlaMetGluL 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    107 CCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAGAGTTGCCATGGAGA 156
```

```
 34 ysIleProValSerAlaPheLeuLeuLeuValAlaLeuSerTyrThrLeu 50
    ||||||||||||||||||||||||||||||||||||||||||||||||||
157 AAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCTCCTACACTCTG 206

 51 AlaArgAspThrThrValLysProGlyAlaLysLysAspThrLysAspSe 67
    ||||||||||||||||||||||||||||||||||||||||||||||||||
207 GCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGACACAAAGGACTC 256

 67 rArgProLysLeuProGlnThrLeuSerArgGlyTrpGlyAspGlnLeuI 84
    ||||||||||||||||||||||||||||||||||||||||||||||||||
257 TCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGGGTGACCAACTCA 306

 84 leTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerLysThrSerAsn 100
    ||||||||||||||||||||||||||||||||||||||||||||||||||
307 TCTGGACTCAGACATATGAAGAAGCTCTATATAAATCCAAGACAAGCAAC 356

101 LysProLeuMetIleIleHisHisLeuAspGluCysProHisSerGlnAl 117
    ||||||||||||||||||||||||||||||||||||||||||||||||||
357 AAACCCTTGATGATTATTCATCACTTGGATGAGTGCCCACACAGTCAAGC 406

117 aLeuLysLysValPheAlaGluAsnLysGluIleGlnLysLeuAlaGluG 134
    ||||||||||||||||||||||||||||||||||||||||||||||||||
407 TTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCAGAAATTGGCAGAGC 456

134 lnPheValLeuLeuAsnLeuValTyrGluThrThrAspLysHisLeuSer 150
    ||||||||||||||||||||||||||||||||||||||||||||||||||
457 AGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTGACAAACACCTTTCT 506

151 ProAspGlyGlnTyrValProArgIleMetPheValAspProSerLeuTh 167
    ||||||||||||||||||||||||||||||||||||||||||||||||||
507 CCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTTGACCCATCTCTGAC 556

167 rValArgAlaAspIleThrGlyArgTyrSerAsnArgLeuTyrAlaTyrG 184
    ||||||||||||||||||||||||||||||||||||||||||||||||||
557 AGTTAGAGCCGATATCACTGGAAGATATTCAAACCGTCTCTATGCTTACG 606

184 luProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLysAlaLeuLys 200
    ||||||||||||||||||||||||||||||||||||||||||||||||||
607 AACCTGCAGATACAGCTCTGTTGCTTGACAACATGAAGAAAGCTCTCAAG 656

201 LeuLeuLysThrGluLeu 206
    |||||||||||||||||
657 TTGCTGAAGACTGAATTG 674
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:AAV29047

seq_documentation_block:

```
ID   AAV29047 standard; cDNA; 866 BP.
XX
AC   AAV29047;
XX
DT   21-AUG-1998  (first entry)
XX
DE   Human protein comprising secretory signal nucleotide sequence 9.
XX
KW   Human protein; secretory signal; nutritional source; cytokine;
KW   immunity;haematopoiesis; activin; inhibin; tumour; chemotactic;
KW   chemokinetic; thrombolytic; anti-inflammatory; inhibition;
KW   stomach cancer cell; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             73..600
FT                   /*tag=  a
FT                   /product= "human protein comprising secretory
FT                            signal"
XX
PN   WO9811217-A2.
XX
PD   19-MAR-1998.
XX
PF   12-SEP-1997;   97WO-JP03239.
XX
PR   13-SEP-1996;   96JP-0243060.
XX
PA   (PROT-) PROTEGENE INC.
PA   (SAGA ) SAGAMI CHEM RES CENTRE.
XX
PI   Kato S,  Kobayashi M,  Sekine S,  Yamaguchi T;
XX
DR   WPI; 1998-207380/18.
DR   P-PSDB; AAW37872.
XX
PT   Human proteins with secretory signal sequences - used to treat
PT   immune deficiencies, infections, tumours, and haematopoietic
PT   disorders, etc.
XX
PS   Claim 4; Pages 114-116; 131pp; English.
XX
CC   This is the nucleotide sequence of a novel human protein comprising
CC   a secretory signal isolated from stomach cancer cells.  Its proteins
CC   can be used as nutritional sources or supplements.  The proteins may
CC   also have cytokine functions, immune modulating functions,
CC   haematopoiesis regulating activity, activin/inhibin regulating
CC   activity, chemotactic/chemokinetic activity, haemostatic and
CC   thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC   activity, tumour inhibition activity.
XX
SQ   Sequence 866 BP; 275 A; 189 C; 178 G; 224 T; 0 other;
```

alignment_scores:

```
            Quality:  195.00              Length:    195
              Ratio:  1.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000
```

alignment_block:

US-09-674-266A-181 x AAV29047 ..

Align seg 1/1 to: AAV29047 from: 1 to: 866

```
 12 GlyProHisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSe 28
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 13 GGACCGCATCCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATC 62

 28 rArgValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValA 45
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 63 CAGAGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGG 112

 45 laLeuSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLys 61
    ||||||||||||||||||||||||||||||||||||||||||||||||||
113 CCCTCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAA 162

 62 LysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGl 78
    ||||||||||||||||||||||||||||||||||||||||||||||||||
163 AAGGACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGG 212

 78 yTrpGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrL 95
    ||||||||||||||||||||||||||||||||||||||||||||||||||
213 TTGGGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATA 262

 95 ysSerLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGlu 111
    ||||||||||||||||||||||||||||||||||||||||||||||||||
263 AATCCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAG 312

112 CysProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIl 128
    ||||||||||||||||||||||||||||||||||||||||||||||||||
313 TGCCCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAAT 362

128 eGlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrT 145
    ||||||||||||||||||||||||||||||||||||||||||||||||||
363 CCAGAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAA 412

145 hrAspLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPhe 161
    |||||||||||||||||||||||||||||||||||||||||||||||||
```

```
     413 CTGACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTT 462

     162 ValAspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAs 178
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     463 GTTGACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAA 512

     178 nArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnM 195
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     513 CCGTCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACA 562

     195 etLysLysAlaLeuLysLeuLeuLysThrGluLeu 206
         |||||||||||||||||||||||||||||||||||
     563 TGAAGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 597

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAC58380

seq_documentation_block:
ID   AAC58380 standard; cDNA; 920 BP.
XX
AC   AAC58380;
XX
DT   29-JAN-2001  (first entry)
XX
DE   Human PRO1030 nucleotide sequence SEQ ID NO:39.
XX
KW   Human; tumour; diagnosis; neoplastic disease; neoplastic cell growth;
KW   proliferation; tumourigenesis; identification; cancer; cytostatic;
KW   nootropic; neuroprotective; antiinflammatory; immunosuppressive;
KW   immunostimulant; antiangiogenic; leukaemia; lymphoid malignancy;
KW   neuronal disorder; glial disorder; astrocytal disorder; angiogenic;
KW   hypothalamic disorder; glandular disorder; macrophagal disorder;
KW   epithelial disorder; stromal disorder; blastocoelic disorder;
KW   inflammatory disorder; immunologic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200053755-A2.
XX
PD   14-SEP-2000.
XX
PF   06-JAN-2000; 2000WO-US00376.
XX
PR   08-MAR-1999;   99WO-US05028.
PR   02-JUN-1999;   99WO-US12252.
PR   23-JUN-1999;   99US-0141037.
PR   07-JUL-1999;   99US-0143048.
PR   26-JUL-1999;   99US-0145698.
PR   30-NOV-1999;   99WO-US28313.
PR   20-DEC-1999;   99WO-US30911.
PR   05-JAN-2000; 2000WO-US00219.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Ashkenazi AJ,  Baker KP,  Goddard A,  Gurney AL,  Hillan KJ,  Roy MA;
PI   Watanabe CK,  Wood WI;
XX
DR   WPI; 2000-572270/53.
DR   P-PSDB; AAB24070.
XX
PT   Thirty PRO polynucleotides encoding PRO polypeptides, useful in the
PT   treatment, diagnosis and prevention of cancer -
XX
PS   Claim 50; Fig 27; 286pp; English.
XX
CC   The present invention describes an isolated antibody  that binds to
CC   one of the human PRO proteins designated PRO212, PRO290, PRO341, PRO535,
CC   PRO619, PRO717, PRO809, PRO830, PRO848, PRO943, PRO1005, PRO1009,
CC   PRO1025, PRO1030, PRO1097, PRO1107, PRO1111, PRO1153, PRO1182, PRO1184,
CC   PRO1187, PRO1281, PRO23, PRO39, PRO834, PRO1317, PRO1710, PRO2094,
CC   PRO2145 OR PRO2198. PRO antagonists can be used to inhibit tumour cell
CC   growth. The PRO polypeptides and nucleotides are useful in the
CC   treatment, diagnosis and prevention of cancer. The antibodies and other
CC   anti-tumour compounds maybe used to treat various conditions, including
CC   those characterised by overexpression and/or activation of the amplified
CC   PRO genes. Exemplary conditions or disorders to be treated with such
CC   antibodies and other compounds include benign or malignant tumours
CC   (e.g., renal, liver, kidney, bladder, breast, gastric, ovarian,
CC   colorectal, prostate, pancreatic, lung, vulva, thyroid, hepatic
CC   carcinomas, sarcomas, glioblastomas, and various head and neck tumours),
CC   leukaemias and lymphoid malignancies, other disorders such as neuronal,
CC   glial, astrocytal, hypothalamic and other glandular, macrophagal,
CC   epithelial, stromal and blastocoelic disorders, and inflammatory,
CC   angiogenic and immunologic disorders. AAC58242 to AAC58366 represent PCR
CC   primers and hybridisation probes used in the isolation of the human PRO
CC   sequences. AAC58367 to AAC58396 and AAB24057 to AAB24089 represent human
CC   PRO polynucleotide and protein sequences given in the exemplification of
CC   the present invention.
XX
SQ   Sequence 920 BP; 296 A; 198 C; 195 G; 231 T; 0 other;


alignment_scores:
            Quality:  195.00              Length:    195
              Ratio:  1.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x AAC58380   ..

 Align seg 1/1  to: AAC58380  from: 1 to: 920

      12 GlyProHisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSe 28
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      88 GGACCGCATCCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATC 137

      28 rArgValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValA 45
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     138 CAGAGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGG 187

      45 laLeuSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLys 61
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     188 CCCTCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAA 237

      62 LysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGl 78
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     238 AAGGACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGG 287

      78 yTrpGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrL 95
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     288 TTGGGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATA 337

      95 ysSerLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGlu 111
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     338 AATCCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAG 387

     112 CysProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIl 128
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     388 TGCCCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAAT 437

     128 eGlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrT 145
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     438 CCAGAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAA 487

     145 hrAspLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPhe 161
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     488 CTGACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTT 537

     162 ValAspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAs 178
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     538 GTTGACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAA 587

     178 nArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnM 195
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     588 TCGTCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACA 637

     195 etLysLysAlaLeuLysLeuLeuLysThrGluLeu 206
```

```
      ||||||||||||||||||||||||||||||||||
  638 TGAAGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 672

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH35026

seq_documentation_block:
ID   AAH35026 standard; cDNA; 943 BP.
XX
AC   AAH35026;
XX
DT   03-SEP-2001  (first entry)
XX
DE   Human colon cancer antigen encoding cDNA SEQ ID NO:2108.
XX
KW   Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW   colorectal carcinoma; ss.
XX
OS   Homo sapiens.
XX
PN   WO200122920-A2.
XX
PD   05-APR-2001.
XX
PF   28-SEP-2000; 2000WO-US26524.
XX
PR   29-SEP-1999;   99US-0157137.
PR   03-NOV-1999;   99US-0163280.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Barash SC,  Birse CE,  Rosen CA;
XX
DR   WPI; 2001-235357/24.
DR   P-PSDB; AAG75621.
XX
PT   Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT   useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS   Claim 1; Page 3593; 9803pp; English.
XX
CC   AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC   cancer-associated nucleic acid molecules (N) and proteins (P), where
CC   the proteins are collectively known as colon cancer antigens. The colon
CC   cancer antigens have cytostatic activity and can be used in gene
CC   therapy and vaccine production. N and P may be used in the prevention,
CC   diagnosis and treatment of diseases associated with inappropriate P
CC   expression. For example, N and P may be used to treat disorders
CC   associated with decreased expression by rectifying mutations or deletions
CC   in a patient's genome that affect the activity of P by expressing
CC   inactive proteins or to supplement the patients own production of P.
CC   Additionally, N may be used to produce the colon cancer-associated Ps,
CC   by inserting the nucleic acids into a host cell and culturing the cell
CC   to express the proteins. N and P can be used in the prevention, diagnosis
CC   and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC   and AAB77789 represent sequences used in the exemplification of the
CC   present invention.
CC   N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC   missing at time of publication, meaning no sequences are present for
CC   SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ   Sequence 943 BP; 311 A; 199 C; 199 G; 229 T; 5 other;


alignment_scores:
           Quality:  194.00              Length:    194
             Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x AAH35026    ..

 Align seg 1/1  to: AAH35026  from: 1  to: 943
```

```
   13 ProHisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerAr 29
      ||||||||||||||||||||||||||||||||||||||||||||||||||
   26 CCGCATCCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAG 75

   29 gValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAlaL 46
      ||||||||||||||||||||||||||||||||||||||||||||||||||
   76 AGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCC 125

   46 euSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLys 62
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  126 TCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAG 175

   63 AspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTr 79
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  176 GACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTG 225

   79 pGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysS 96
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  226 GGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAAT 275

   96 erLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCys 112
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  276 CCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGC 325

  113 ProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGl 129
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  326 CCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCA 375

  129 nLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrA 146
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  376 GAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTG 425

  146 spLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheVal 162
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  426 ACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTT 475

  163 AspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnAr 179
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  476 GACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAAYCG 525

  179 gLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetL 196
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  526 TCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACATGA 575

  196 ysLysAlaLeuLysLeuLeuLysThrGluLeu 206
      |||||||||||||||||||||||||||||||
  576 AGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 607

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAZ77486

seq_documentation_block:
ID   AAZ77486 standard; cDNA; 1018 BP.
XX
AC   AAZ77486;
XX
DT   10-APR-2000  (first entry)
XX
DE   Human ovarian tumor cDNA library derived EST fragment 37.
XX
KW   Expressed sequence tag; EST; human; ovarian tumor; anticancer;
KW   gene therapy; treatment; ss.
XX
OS   Homo sapiens.
XX
PN   DE19817557-A1.
XX
PD   21-OCT-1999.
XX
PF   09-APR-1998;   98DE-1017557.
XX
PR   09-APR-1998;   98DE-1017557.
XX
PA   (META-) METAGEN GES GENOMFORSCHUNG MBH.
```

```
XX
PI   Rosenthal A,  Specht T,  Hinzmann B,  Schmitt A,  Pilarsky C,  Dahl E;
XX
DR   WPI; 1999-591920/51.
DR   P-PSDB; AAY76590.
XX
PT   New nucleic acid sequences expressed in ovarian, and some other, cancer
PT   tissues, and derived polypeptides, for treatment of ovarian cancer and
PT   identification of therapeutic agents  -
XX
PS   Claim 3; Page 167; 310pp; German.
XX
CC   This invention describes novel nucleic acid (cDNA) sequences (A) which
CC   have anticancer activity and are highly expressed in ovarian tumor
CC   tissue (and some also in testis and breast cancer tissue). The products
CC   of the invention can be used for gene therapy. (A) are used (i) for
CC   recombinant expression of polypeptides (B) and (ii) to isolate complete
CC   genes. (B) are used (i) to identify agents suitable for treatment of
CC   ovarian cancer; (ii) directly for treating this form of cancer
CC   (including expression from gene therapy vectors) and (iii) for generation
CC   of specific antibodies. (A) are identified by assembling ESTs (expressed
CC   sequence tags) from a particular tissue type before comparison of
CC   expression patterns. This allows a significantly longer fragment of the
CC   gene to be revealed, so should reduce the number of failures associated
CC   with the fact that ESTs from different libraries may represent different
CC   parts of the same unknown gene, distorting the estimated frequency of
CC   occurrence in a particular tissue. AAZ77450-Z77572 represent the human
CC   ovarian tumor cDNA library derived EST fragments described in the method
CC   of the invention and encode the protein fragments represented in
CC   AAY76505-Y76638.
XX
SQ   Sequence 1018 BP; 343 A; 212 C; 216 G; 247 T; 0 other;


alignment_scores:
            Quality:  194.00              Length:    194
              Ratio:  1.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x AAZ77486   ..

 Align seg 1/1  to: AAZ77486  from: 1  to: 1018

      13 ProHisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerAr 29
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      91 CCGCATCCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAG 140

      29 gValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAlaL 46
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     141 AGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCC 190

      46 euSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLys 62
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     191 TCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAG 240

      63 AspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTr 79
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     241 GACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTG 290

      79 pGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysS 96
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     291 GGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAAT 340

      96 erLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCys 112
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     341 CCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGC 390

     113 ProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGl 129
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     391 CCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCA 440

     129 nLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrA 146
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     441 GAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTG 490

     146 spLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheVal 162
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     491 ACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTT 540

     163 AspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnAr 179
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     541 GACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAACCG 590

     179 gLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetL 196
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     591 TCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACATGA 640

     196 ysLysAlaLeuLysLeuLeuLysThrGluLeu 206
         |||||||||||||||||||||||||||||||
     641 AGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 672

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:AAV19155

seq_documentation_block:
ID   AAV19155 standard; cDNA; 875 BP.
XX
AC   AAV19155;
XX
DT   28-AUG-1998  (first entry)
XX
DE   Human XAG growth factor huXAG-1 cDNA.
XX
KW   HuXAG-1; XAG; growth factor; colon cancer-specific gene;
KW   tumour marker; breast disease; liver disease; lung disease;
KW   emphysema; wound healing; diagnosis; therapy; human; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             71..598
FT                   /*tag=  a
FT   sig_peptide     71..130
FT                   /*tag=  b
FT   mat_peptide     131..595
FT                   /*tag=  c
XX
PN   WO9807749-A1.
XX
PD   26-FEB-1998.
XX
PF   22-AUG-1997;   97WO-US14139.
XX
PR   23-AUG-1996;   96WO-US13766.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Dillon PJ,  Ebner R,  Endress GA,  Yu G;
XX
DR   WPI; 1998-169093/15.
DR   P-PSDB; AAW37844.
XX
PT   New isolated human XAG growth factor(s) - used to develop products
PT   for treating e.g. liver, lung or breast diseases or
PT   hyperproliferative disorders, e.g. cancer.
XX
PS   Claim 5; Fig 1; 141pp; English.
XX
CC   This cDNA clone codes for huXAG-1 (see AAW37844), a member of a novel
CC   family of human growth factors also including huXAG-2 (see AAW37845)
CC   and huXAG-3 (see AAW37846).  These proteins share homology with the
CC   XAG protein of Xenopus laevis, which is involved in embryogenesis
CC   and is expressed in adult tissue.  Expression of huXAG-1 has been
CC   discovered in colon cancer tissue, with no corresponding expression
CC   in healthy colon tissue.  The huXAG-1 gene, also designated
CC   colon cancer-specific gene (CCSG), provides a molecular marker for
```

```
CC   colon cancer.  huXAG-1 cDNA was isolated from a cDNA library
CC   derived from human colon cancer tissue.  Vectors, host cells,
CC   antibodies, and screening methods for identifying agonists and
CC   antagonists of huXAG-1 are provided.  HuXAG polypeptides are growth
CC   factors and can be used to stimulate proliferation of cells.  They
CC   can be used to stimulate the proliferation and differentiation of
CC   hepatocytes to alleviate or treat liver diseases and pathologies
CC   such as fulminant liver failure caused by cirrhosis, liver damage
CC   caused by viral hepatitis and toxic substances.  They can also be
CC   used to stimulate or promote liver regeneration, e.g. after
CC   surgery.  They can also be used to prevent and heal damage to the
CC   lungs caused by various pathological states.  They can be used to
CC   stimulate proliferation and differentiation and promote the repair
CC   of alveoli and bronchiolar epithelium to prevent, attenuate, or
CC   treat acute or chronic lung damage, e.g. emphysema, which results
CC   in the progressive loss of alveoli, and inhalation injuries, e.g.
CC   resulting from smoke inhalation and burns, that cause necrosis of
CC   the bronchiolar epithelium and alveoli.  They can also be used to
CC   stimulate the proliferation and differentiation of breast tissue
CC   and could therefore be used to promote healing of breast tissue
CC   injury due to surgery, trauma or cancer.  Antagonists can be used
CC   to treat hyperproliferative disorders, including cancer, in
CC   particular hepatocellular carcinoma, osteoclastoma, breast cancer,
CC   or colon cancer.  The products can also be used for detection and
CC   diagnosis.
XX
SQ   Sequence 875 BP; 283 A; 189 C; 180 G; 223 T; 0 other;


alignment_scores:
           Quality:  192.00            Length:    192
             Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x AAV19155   ..

 Align seg 1/1  to: AAV19155  from: 1  to: 875

      15 ProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerArgValAl 31
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      20 CCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAGAGTTGC 69

      31 aMetGluLysIleProValSerAlaPheLeuLeuLeuValAlaLeuSerT 48
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      70 CATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCTCCT 119

      48 yrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLysAspThr 64
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     120 ACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGACACA 169

      65 LysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTrpGlyAs 81
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     170 AAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGGGTGA 219

      81 pGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerLysT 98
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     220 CCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAATCCAAGA 269

      98 hrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCysProHis 114
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     270 CAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGCCCACAC 319

     115 SerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGlnLysLe 131
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     320 AGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCAGAAATT 369

     131 uAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrAspLysH 148
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     370 GGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTGACAAAC 419

     148 isLeuSerProAspGlyGlnTyrValProArgIleMetPheValAspPro 164
         ||||||||||||||||||||||||||||||||||||||||||||||||||

     420 ACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTTGACCCA 469

     165 SerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnArgLeuTy 181
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     470 TCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAATCGTCTCTA 519

     181 rAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLysA 198
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     520 TGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACATGAAGAAAG 569

     198 laLeuLysLeuLeuLysThrGluLeu 206
         |||||||||||||||||||||||||
     570 CTCTCAAGTTGCTGAAGACTGAATTG 595

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAF63314

seq_documentation_block:
ID   AAF63314 standard; cDNA; 875 BP.
XX
AC   AAF63314;
XX
DT   04-MAY-2001  (first entry)
XX
DE   Human huXAG-1/CCSG colon cancer specific gene cDNA.
XX
KW   Human; growth factor; huXAG-1; colon cancer specific gene; CCSG;
KW   cell proliferation; liver disease; fulminant liver failure; cirrhosis;
KW   hepatitis; cancer; colon cancer; colorectal carcinoma; ss.
XX
OS   Homo sapiens.
XX
PN   US6171816-B1.
XX
PD   09-JAN-2001.
XX
PF   22-AUG-1997;   97US-0916576.
XX
PR   23-AUG-1996;   96US-0024347.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Yu G,  Dillon PJ,  Ebner R,  Endress GA;
XX
DR   WPI; 2001-136872/14.
DR   P-PSDB; AAB72203.
XX
PT   Novel human growth factor polypeptide useful for diagnosing and
PT   treating colon cancer and liver diseases, to prevent and heal damage to
PT   the lungs and for identifying modulators of therapeutic use -
XX
PS   Claim 45; Fig 1; 59pp; English.
XX
CC   This invention relates to a human growth factor polypeptide huXAG-1 also
CC   known as a colon cancer specific gene (CCSG). HuXAG-1 stimulates cell
CC   proliferation as a growth factor. The HuXAG-1 protein is useful for
CC   identifying compounds capable of enhancing or inhibiting cellular
CC   response induced by huXAG-1. The protein is also useful for stimulating
CC   proliferation of cells e.g. colon, breast, liver and lung cells, and
CC   hepatocytes. It is useful for alleviating or treating liver diseases and
CC   pathologies such as fulminant liver failure caused by cirrhosis, liver
CC   damage caused by viral hepatitis and toxic substances, for preventing and
CC   treating damage to the lungs caused by various pathological states and
CC   for promoting healing of breast tissue injury due to surgery, trauma or
CC   cancer. HuXAG-1 and the identified antagonist are useful for treating
CC   cancer, in particular colon cancer. Detecting altered levels of huXAG-1
CC   and its polynucleotides are useful for diagnosing or detecting cancer in
CC   mammals. The gene encoding huXAG-1 is useful for monitoring human
CC   colorectal carcinoma. huXAG-1 nucleic acid molecules are also useful for
CC   chromosome identification. The present sequence represents cDNA encoding
CC   huXAG-1.
XX
SQ   Sequence 875 BP; 283 A; 189 C; 180 G; 223 T; 0 other;
```

```
alignment_scores:
            Quality:  192.00              Length:    192
              Ratio:  1.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x AAF63314   ..

 Align seg 1/1  to: AAF63314  from: 1  to: 875

      15 ProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerArgValAl 31
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      20 CCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAGAGTTGC 69

      31 aMetGluLysIleProValSerAlaPheLeuLeuLeuValAlaLeuSerT 48
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      70 CATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCTCCT 119

      48 yrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLysAspThr 64
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     120 ACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGACACA 169

      65 LysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTrpGlyAs 81
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     170 AAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGGGTGA 219

      81 pGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerLysT 98
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     220 CCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAATCCAAGA 269

      98 hrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCysProHis 114
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     270 CAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGCCCACAC 319

     115 SerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGlnLysLe 131
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     320 AGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCAGAAATT 369

     131 uAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrAspLysH 148
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     370 GGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTGACAAAC 419

     148 isLeuSerProAspGlyGlnTyrValProArgIleMetPheValAspPro 164
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     420 ACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTTGACCCA 469

     165 SerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnArgLeuTy 181
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     470 TCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAATCGTCTCTA 519

     181 rAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLysA 198
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     520 TGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACATGAAGAAAG 569

     198 laLeuLysLeuLeuLysThrGluLeu 206
         |||||||||||||||||||||||||
     570 CTCTCAAGTTGCTGAAGACTGAATTG 595

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:AAV59320

seq_documentation_block:
ID   AAV59320 standard; cDNA; 881 BP.
XX
AC   AAV59320;
XX
DT   21-DEC-1998  (first entry)
XX
DE   Nucleotide sequence encoding zsig10 polypeptide.
XX
KW   ss; human; mucous-mediated function; adhesion; tumour metastasis;
KW   bacterial colonisation; microbial infection; AIDS; cystic fibrosis;
KW   chronic obstructive pulmonary disease; asthma; Crohn's disease;
KW   sinonasal inflammatory disease; inflammatory bowel disease; bronchitis.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             63..590
FT                   /*tag= a
FT                   /product= "zsig10"
XX
PN   WO9841627-A1.
XX
PD   24-SEP-1998.
XX
PF   18-MAR-1998;   98WO-US05251.
XX
PR   19-MAR-1997;   97US-0039631.
XX
PA   (ZYMO ) ZYMOGENETICS INC.
XX
PI   Sheppard PO;
XX
DR   WPI; 1998-531566/45.
DR   P-PSDB; AAW77365.
XX
PT   New isolated mucous-associated polypeptide, zsig10 - used to develop
PT   products for treating e.g. tumour metastasis, microbial infections,
PT   cystic fibrosis, asthma, bronchitis or inflammatory bowel disease
XX
PS   Claim 37; Page 80-81; 109pp; English.
XX
CC   The human polypeptide zsig10 is involved in mucous-mediated functions
CC   such as adhesion.  The products of the invention can be used in the study
CC   and treatment of e.g. tumour metastasis, bacterial colonisation,
CC   susceptibility to and persistence of infection, microbial infections,
CC   AIDS, cystic fibrosis, chronic obstructive pulmonary disease, asthma,
CC   sinonasal inflammatory disease, inflammatory bowel disease, bronchitis,
CC   or Crohn's disease.  The products can also be used for detection,
CC   diagnosis and drug screening.
XX
SQ   Sequence 881 BP; 293 A; 185 C; 180 G; 223 T; 0 other;


alignment_scores:
            Quality:  191.00              Length:    191
              Ratio:  1.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x AAV59320   ..

 Align seg 1/1  to: AAV59320  from: 1  to: 881

      16 SerArgArgLeuThrGlnGlyArgTrpValArgLysSerArgValAlaMe 32
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      15 AGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAGAGTTGCCAT 64

      32 tGluLysIleProValSerAlaPheLeuLeuLeuValAlaLeuSerTyrT 49
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      65 GGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCTCCTACA 114

      49 hrLeuAlaArgAspThrThrValLysProGlyAlaLysLysAspThrLys 65
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     115 CTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGACACAAAG 164

      66 AspSerArgProLysLeuProGlnThrLeuSerArgGlyTrpGlyAspGl 82
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     165 GACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGGGTGACCA 214

      82 nLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerLysThrS 99
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     215 ACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAATCCAAGACAA 264

      99 erAsnLysProLeuMetIleIleHisHisLeuAspGluCysProHisSer 115
```

```
     ||||||||||||||||||||||||||||||||||||||||||||||||||
265  GCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGCCCACACAGT 314

116  GlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGlnLysLeuAl 132
     ||||||||||||||||||||||||||||||||||||||||||||||||||
315  CAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCAGAAATTGGC 364

132  aGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrAspLysHisL 149
     ||||||||||||||||||||||||||||||||||||||||||||||||||
365  AGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTGACAAACACC 414

149  euSerProAspGlyGlnTyrValProArgIleMetPheValAspProSer 165
     ||||||||||||||||||||||||||||||||||||||||||||||||||
415  TTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTTGACCCATCT 464

166  LeuThrValArgAlaAspIleThrGlyArgTyrSerAsnArgLeuTyrAl 182
     ||||||||||||||||||||||||||||||||||||||||||||||||||
465  CTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAATCGTCTCTATGC 514

182  aTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLysAlaL 199
     ||||||||||||||||||||||||||||||||||||||||||||||||||
515  TTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACATGAAGAAAGCTC 564

199  euLysLeuLeuLysThrGluLeu 206
     ||||||||||||||||||||||
565  TCAAGTTGCTGAAGACTGAATTG 587
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAS13480

seq_documentation_block:

```
ID   AAS13480 standard; DNA; 543 BP.
XX
AC   AAS13480;
XX
DT   19-DEC-2001  (first entry)
XX
DE   DNA encoding breast cancer cell membrane protein 7 (BCMP 7).
XX
KW   Breast cancer cell membrane protein 7; BCMP 7; breast cancer;
KW   metastasis; cytostatic.; gene therapy; antibody; antisense;
KW   vaccine; chromosome 7p21.3; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             11..538
FT                   /*tag=  a
FT                   /product= "BCMP 7"
FT                   /note= "Breast cancer cell membrane protein 7"
FT   sig_peptide     11..70
FT                   /*tag=  b
FT   mat_peptide     71..535
FT                   /*tag=  c
FT                   /note= "Mature BCMP 7"
XX
PN   WO200163290-A1.
XX
PD   30-AUG-2001.
XX
PF   21-FEB-2001; 2001WO-GB00734.
XX
PR   25-FEB-2000; 2000GB-0004576.
XX
PA   (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI   Boyd RS,  Stamps AC,  Terrett JA,  Tyson KL;
XX
DR   WPI; 2001-570651/64.
DR   P-PSDB; AAU08804.
XX
PT   Diagnosing, preventing and treating breast cancer using a breast cancer
PT   cell membrane protein BCMP 7 -
XX
PS   Claim 6; Fig 1; 62pp; English.
XX
CC   The invention describes the novel use of a protein found in breast
CC   cancer cell membranes (BCMP 7) for diagnosing, preventing and treating
CC   breast cancers. The peptide has cytostatic action and potential uses in
CC   gene therapy and in vaccines. The polypeptide, antisense nucleic acids,
CC   or fusion proteins comprising and Green Fluorescent protein or the DsRed
CC   Fluorescent protein, antibodies specific for and/or nucleic acid are
CC   used for the prevention and/or treatment of breast cancer. Antibodies
CC   against may also be used for screening for and/or diagnosis of breast
CC   cancer in a patient. The method for monitoring/assessing breast cancer
CC   treatment in a patient and for the identification of metastatic breast
CC   cancer cells in samples from a patient. This sequence encodes breast
CC   cancer cell associated protein 7 (BCMP 7), located on chromosome
CC   7p21.3, described in the method of the invention.
XX
SQ   Sequence 543 BP; 170 A; 135 C; 107 G; 131 T; 0 other;
```

alignment_scores:

```
            Quality:  178.00              Length:    178
              Ratio:  1.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000
```

alignment_block:

US-09-674-266A-181 x AAS13480 ..

Align seg 1/1 to: AAS13480 from: 1 to: 543

```
 29  ArgValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAl 45
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  2  AGAGTTGCCATGGAGAAAATCCCAGTGTCAGCATTCTTGCTCCTTGTGGC 51

 45  aLeuSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysL 62
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 52  CCTCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAA 101

 62  ysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGly 78
     ||||||||||||||||||||||||||||||||||||||||||||||||||
102  AGGACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGT 151

 79  TrpGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLy 95
     ||||||||||||||||||||||||||||||||||||||||||||||||||
152  TGGGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAA 201

 95  sSerLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluC 112
     ||||||||||||||||||||||||||||||||||||||||||||||||||
202  ATCCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGT 251

112  ysProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIle 128
     ||||||||||||||||||||||||||||||||||||||||||||||||||
252  GCCCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATC 301

129  GlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrTh 145
     ||||||||||||||||||||||||||||||||||||||||||||||||||
302  CAGAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAAC 351

145  rAspLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheV 162
     ||||||||||||||||||||||||||||||||||||||||||||||||||
352  TGACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTG 401

162  alAspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsn 178
     ||||||||||||||||||||||||||||||||||||||||||||||||||
402  TTGACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAAC 451

179  ArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMe 195
     ||||||||||||||||||||||||||||||||||||||||||||||||||
452  CGTCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACAT 501

195  tLysLysAlaLeuLysLeuLeuLysThrGluLeu 206
     ||||||||||||||||||||||||||||||||||
502  GAAGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 535
```

```
seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAZ40803

seq_documentation_block:
ID   AAZ40803 standard; DNA; 1689 BP.
XX
AC   AAZ40803;
XX
DT   18-JAN-2000  (first entry)
XX
DE   Secreted protein EST coding sequence 108-008-5-0-A6-FL.
XX
KW   Secreted protein; fingerprint identification technique;
KW   chromosome mapping; human; hereditary disease; diagnosis; cancer;
KW   hyperlipidaemia; cardiovascular; neurodegenerative disorder; therapy;
KW   autoimmune disease; rheumatic disease; embryogenic disorder; myopathy;
KW   renal injury; amino aciduria; hypoglycaemia; male rat infertility;
KW   hypertension; ss.
XX
OS   Homo sapiens.
XX
PN   WO9940189-A2.
XX
PD   12-AUG-1999.
XX
PF   09-FEB-1999;   99WO-IB00282.
XX
PR   09-FEB-1998;   98US-0074121.
PR   13-APR-1998;   98US-0081563.
PR   10-AUG-1998;   98US-0096116.
PR   04-SEP-1998;   98US-0099273.
XX
PA   (GEST ) GENSET.
XX
PI   Bougueleret L,  Duclert A,  Dumas Milne Edwards J;
XX
DR   WPI; 1999-600966/51.
DR   P-PSDB; AAY59675.
XX
PT   Extended cDNAs useful for expressing secreted proteins and to obtain
PT   specific antibodies -
XX
PS   Claim 1; Page 168-169; 244pp; English.
XX
CC   This sequence encodes a human secreted protein of the invention. The
CC   extended cDNAs (or genomic DNAs obtainable from them) may be used to
CC   prepare PCR primers and probes. These are useful for forensic matching or
CC   positive identification by DNA sequencing. They may also be used in
CC   alternative fingerprint identification techniques. Antibodies against the
CC   proteins encoded by the extended cDNAs are useful in identification of
CC   tissue types or cell species, as well as identifying tissue specific
CC   soluble proteins. The sequences can be used for chromosome mapping and
CC   identification of genes associated with hereditary diseases or drug
CC   response. signal sequences from the cDNAs can be used in construction of
CC   secretion vectors. Other sequences derived from the extended cDNAs can be
CC   used to clone upstream genomic DNA sequences including promoters. This is
CC   in turn useful for identifying proteins that interact with promoter
CC   sequences. Some of the proteins may be useful in diagnosing and treating
CC   several disorders including, but not limited to: cancer, hyperlipidaemia,
CC   cardiovascular and neurodegenerative disorders, autoimmune diseases, and
CC   rheumatic diseases, embryogenic disorders, hypertension, renal injury,
CC   amino acidurias, hypoglycaemia, male rat infertility and myopathies.
XX
SQ   Sequence 1689 BP; 552 A; 350 C; 335 G; 452 T; 0 other;


alignment_scores:
            Quality:  178.00              Length:    178
              Ratio:  1.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x AAZ40803   ..

 Align seg 1/1  to: AAZ40803  from: 1  to: 1689

      29 ArgValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAl 45
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      42 AGAGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGC 91

      45 aLeuSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysL 62
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      92 CCTCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAA 141

      62 ysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGly 78
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     142 AGGACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGT 191

      79 TrpGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLy 95
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     192 TGGGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAA 241

      95 sSerLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluC 112
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     242 ATCCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGT 291

     112 ysProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIle 128
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     292 GCCCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATC 341

     129 GlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrTh 145
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     342 CAGAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAAC 391

     145 rAspLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheV 162
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     392 TGACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTG 441

     162 alAspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsn 178
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     442 TTGACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAAT 491

     179 ArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMe 195
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     492 CGTCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACAT 541

     195 tLysLysAlaLeuLysLeuLeuLysThrGluLeu 206
         |||||||||||||||||||||||||||||||||
     542 GAAGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 575

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:AAV29048

seq_documentation_block:
ID   AAV29048 standard; cDNA; 525 BP.
XX
AC   AAV29048;
XX
DT   21-AUG-1998  (first entry)
XX
DE   Open reading frame human protein comprising secretory signal 9.
XX
KW   Human protein; secretory signal; nutritional source; cytokine;
KW   immunity;haematopoiesis; activin; inhibin; tumour; chemotactic;
KW   chemokinetic; thrombolytic; anti-inflammatory; inhibition;
KW   stomach cancer cell; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..525
FT                   /*tag=  a
FT                   /product= "human protein comprising secretory
FT                            signal"
XX
PN   WO9811217-A2.
XX
PD   19-MAR-1998.
```

```
XX
PF   12-SEP-1997;   97WO-JP03239.
XX
PR   13-SEP-1996;   96JP-0243060.
XX
PA   (PROT-) PROTEGENE INC.
PA   (SAGA ) SAGAMI CHEM RES CENTRE.
XX
PI   Kato S,  Kobayashi M,  Sekine S,  Yamaguchi T;
XX
DR   WPI; 1998-207380/18.
DR   P-PSDB; AAW37872.
XX
PT   Human proteins with secretory signal sequences - used to treat
PT   immune deficiencies, infections, tumours, and haematopoietic
PT   disorders, etc.
XX
PS   Claim 3; Pages 88; 131pp; English.
XX
CC   This is the nucleotide sequence of the open reading frame of a novel
CC   human protein comprising a secretory signal (AAV29047), isolated from
CC   stomach cancer cells.  Its proteins can be used as nutritional sources
CC   or supplements.  The proteins may also have cytokine functions,
CC   immune modulating  functions, haematopoiesis regulating activity,
CC   activin/inhibin regulating activity, chemotactic/chemokinetic activity,
CC   haemostatic and thrombolytic activity, receptor/ligand activity,
CC   anti-inflammatory activity, tumour inhibition activity.
XX
SQ   Sequence 525 BP; 162 A; 131 C; 103 G; 129 T; 0 other;


alignment_scores:
            Quality:  175.00              Length:    175
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x AAV29048    ..

 Align seg 1/1  to: AAV29048  from: 1  to: 525

      32 MetGluLysIleProValSerAlaPheLeuLeuLeuValAlaLeuSerTy 48
         |||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCTCCTA 50

      48 rThrLeuAlaArgAspThrThrValLysProGlyAlaLysLysAspThrL 65
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      51 CACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGACACAA 100

      65 ysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTrpGlyAsp 81
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     101 AGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGGGTGAC 150

      82 GlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerLysTh 98
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     151 CAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAATCCAAGAC 200

      98 rSerAsnLysProLeuMetIleIleHisHisLeuAspGluCysProHisS 115
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     201 AAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGCCCACACA 250

     115 erGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGlnLysLeu 131
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     251 GTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCAGAAATTG 300

     132 AlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrAspLysHi 148
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     301 GCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTGACAAACA 350

     148 sLeuSerProAspGlyGlnTyrValProArgIleMetPheValAspProS 165
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     351 CCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTTGACCCAT 400

     165 erLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnArgLeuTyr 181
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     401 CTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAACCGTCTCTAT 450

     182 AlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLysAl 198
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     451 GCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACATGAAGAAAGC 500

     198 aLeuLysLeuLeuLysThrGluLeu 206
         |||||||||||||||||||||||||
     501 TCTCAAGTTGCTGAAGACTGAATTG 525

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAC00115

seq_documentation_block:
ID   AAC00115 standard; cDNA; 468 BP.
XX
AC   AAC00115;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 113.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   P-PSDB; AAG00109.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 113; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. An ORF has been identified within the
CC   sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC   derived from 30 different tissues. EST sequences usually correspond
CC   mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC   often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC   well suited for isolating cDNA sequences derived from the 5' ends of
CC   mRNAs and even in those cases where longer cDNA sequences have been
CC   obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC   mRNAs with intact 5' ends and can therefore be used to obtain full length
CC   cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC   gene therapy and chromosome mapping procedures. They are used to obtain
CC   upstream regulatory sequences and to design expression and secretion
CC   vectors.
XX
SQ   Sequence 468 BP; 142 A; 122 C; 98 G; 106 T; 0 other;


alignment_scores:
            Quality:  146.00              Length:    146
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
```

```
US-09-674-266A-181 x AAC00115   ..

Align seg 1/1  to: AAC00115  from: 1  to: 468

     12 GlyProHisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSe 28
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     13 GGACCGCATCCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATC 62

     28 rArgValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValA 45
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     63 CAGAGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGG 112

     45 laLeuSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLys 61
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    113 CCCTCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAA 162

     62 LysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGl 78
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    163 AAGGACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGG 212

     78 yTrpGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrL 95
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    213 TTGGGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATA 262

     95 ysSerLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGlu 111
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    263 AATCCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAG 312

    112 CysProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIl 128
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    313 TGCCCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAAT 362

    128 eGlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrT 145
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    363 CCAGAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAA 412

    145 hrAspLysHisLeuSerProAspGlyGlnTyrValPro 157
        |||||||||||||||||||||||||||||||||||||
    413 CTGACAAACACCTTTCTCCTGATGGCCAGTATGTCCCA 450

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAF44884

seq_documentation_block:
ID   AAF44884 standard; cDNA; 778 BP.
XX
AC   AAF44884;
XX
DT   28-MAR-2001  (first entry)
XX
DE   Human breast cancer related protein coding sequence SEQ ID NO: 40.
XX
KW   Human; breast cancer; diagnosis; therapy; vaccine; ss.
XX
OS   Homo sapiens.
XX
PN   WO200078960-A2.
XX
PD   28-DEC-2000.
XX
PF   23-JUN-2000; 2000WO-US17536.
XX
PR   23-JUN-1999;   99US-0140903.
PR   12-OCT-1999;   99US-0158980.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Yuqiu J,  Mitcham JL;
XX
DR   WPI; 2001-041426/05.
XX
PT   New polynucleotides encoding breast tumor specific proteins, useful for
PT   prevention, treatment and diagnosis of breast cancer -
XX
PS   Claim 25; Page 132; 165pp; English.
XX
CC   The present invention provides the coding sequences for a number of
CC   breast cancer related proteins. These can be used in vaccinations
CC   against, diagnosis of and treatment of cancer, particularly breast
CC   cancer.
XX
SQ   Sequence 778 BP; 245 A; 175 C; 158 G; 193 T; 7 other;


alignment_scores:
            Quality:  134.00              Length:    134
              Ratio:  1.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x AAF44884   ..

 Align seg 1/1  to: AAF44884  from: 1  to: 778

      54 ThrThrValLysProGlyAlaLysLysAspThrLysAspSerArgProLy 70
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     125 ACCACAGTCAAACCTGGAGCCAAAAAGGACACAAAGGACTCTCGACCCAA 174

      70 sLeuProGlnThrLeuSerArgGlyTrpGlyAspGlnLeuIleTrpThrG 87
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     175 ACTGCCCCAGACCCTCTCCAGAGGTTGGGGTGACCAACTCATCTGGACTC 224

      87 lnThrTyrGluGluAlaLeuTyrLysSerLysThrSerAsnLysProLeu 103
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     225 AGACATATGAAGAAGCTCTATATAAATCCAAGACAAGCAACAAACCCTTG 274

     104 MetIleIleHisHisLeuAspGluCysProHisSerGlnAlaLeuLysLy 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     275 ATGATTATTCATCACTTGGATGAGTGCCCACACAGTCAAGCTTTAAAGAA 324

     120 sValPheAlaGluAsnLysGluIleGlnLysLeuAlaGluGlnPheValL 137
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     325 AGTGTTTGCTGAAAATAAAGAAATCCAGAAATTGGCAGAGCAGTTTGTCC 374

     137 euLeuAsnLeuValTyrGluThrThrAspLysHisLeuSerProAspGly 153
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     375 TCCTCAATCTGGTTTATGAAACAACTGACAAACACCTTTCTCCTGATGGC 424

     154 GlnTyrValProArgIleMetPheValAspProSerLeuThrValArgAl 170
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     425 CAGTATGTCCCCAGGATTATGTTTGTTGACCCATCTCTGACAGTTAGAGC 474

     170 aAspIleThrGlyArgTyrSerAsnArgLeuTyrAlaTyrGluProAlaA 187
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     475 CGATATCACTGGAAGATATTCAAATCGTCTCTATGCTTACGAACCTGCAG 524

     187 sp 187
         ||
     525 AT 526

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAA77903

seq_documentation_block:
ID   AAA77903 standard; cDNA; 409 BP.
XX
AC   AAA77903;
XX
DT   14-NOV-2000  (first entry)
XX
DE   cDNA encoding human colon tumour polypeptide, SEQ ID NO:183.
XX
KW   Human colon tumour polypeptide; tumour antigen; cancer; vaccine;
KW   immunotherapy; diagnosis; progression; ss.
XX
OS   Homo sapiens.
XX
PN   WO200037643-A2.
```

```
XX
PD   29-JUN-2000.
XX
PF   23-DEC-1999;   99WO-US30909.
XX
PR   23-DEC-1998;   98US-0221298.
PR   02-JUL-1999;   99US-0347496.
PR   22-SEP-1999;   99US-0401064.
PR   19-NOV-1999;   99US-0444242.
PR   02-DEC-1999;   99US-0454150.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J,  Lodes MJ,  Secrist H,  Benson DR,  Meagher MJ,  Stolk J;
PI   Wang T,  Yuqiu J;
XX
DR   WPI; 2000-442671/38.
XX
PT   New colon tumor polypeptides used to inhibit the development of cancer,
PT   especially colon cancer, and for diagnosing and monitoring the
PT   progression of the cancer -
XX
PS   Claim 29; Page 144; 229pp; English.
XX
CC   Sequences AAA77722-A78199 represent 478 cDNAs encoding proteins or
CC   portions of proteins which are associated with human colon tumours.
CC   The invention also specifically discloses 8 human colon tumour proteins
CC   (AAB11897-B11904). The nucleic acids, the polypeptides they encode, and
CC   antigen presenting cells (APCs, preferably dendritic cells) expressing
CC   such polypeptides may be used in vaccines that target tumour cells,
CC   especially colon tumour cells, thereby inhibiting the development of
CC   cancer. T-cells specific for the polypeptide expressed by the APC are
CC   used to remove tumour cells from biological samples, especially blood or
CC   fractions thereof. The sample or the isolated T-cells specific for the
CC   polypeptide can then be used to inhibit cancer development. CD4+ and/or
CC   CD8+ T-cells from a patient may be incubated with a polypeptide or
CC   nucleic acid of the invention, or an APC expressing such a polypeptide,
CC   to cause the proliferation of specific T-cells. The T-cells can be
CC   cloned and then administered back to the patient to inhibit cancer
CC   development. Nucleic acids encoding the polypeptides and antibodies
CC   against the polypeptides may be used to determine the expression level
CC   of a tumour protein of the invention, and therefore to determine whether
CC   cancer cells are present. Such diagnostic methods may also be used to
CC   monitor the progression of a cancer by repeating the processes at time
CC   intervals, and comparing the current result to previous results. The
CC   present sequence represents a cDNA encoding a human colon tumour
CC   polypeptide.
XX
SQ   Sequence 409 BP; 131 A; 104 C; 80 G; 94 T; 0 other;


alignment_scores:
           Quality:  133.00              Length:    133
             Ratio:  1.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x AAA77903   ..

 Align seg 1/1  to: AAA77903  from: 1  to: 409

      48 TyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLysAspTh 64
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      11 TACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGACAC 60

      64 rLysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTrpGlyA 81
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      61 AAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGGGTG 110

      81 spGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerLys 97
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     111 ACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAATCCAAG 160

      98 ThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCysProHi 114
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     161 ACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGCCCACA 210

     114 sSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGlnLysL 131
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     211 CAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCAGAAAT 260

     131 euAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrAspLys 147
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     261 TGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTGACAAA 310

     148 HisLeuSerProAspGlyGlnTyrValProArgIleMetPheValAspPr 164
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     311 CACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTTGACCC 360

     164 oSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnArgLeu 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     361 ATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAATCGTCTC 409

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAI28641

seq_documentation_block:
ID   AAI28641 standard; cDNA; 409 BP.
XX
AC   AAI28641;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Colon tumour related determined cDNA sequence for CT140.
XX
KW   Human; immunotherapy; diagnosis; colon cancer; colon tumour;
KW   immunogenic; gene therapy; vaccine; colonic cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200149716-A2.
XX
PD   12-JUL-2001.
XX
PF   29-DEC-2000; 2000WO-US35596.
XX
PR   30-DEC-1999;   99US-0476296.
PR   10-JAN-2000; 2000US-0480321.
PR   15-FEB-2000; 2000US-0504629.
PR   06-MAR-2000; 2000US-0519444.
PR   19-MAY-2000; 2000US-0575251.
PR   29-JUN-2000; 2000US-0609448.
PR   28-AUG-2000; 2000US-0649811.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J,  Lodes MJ,  Secrist H,  Benson DR,  Meagher MJ,  Stolk JA;
PI   King GE,  Wang T,  Jiang Y;
XX
DR   WPI; 2001-441847/47.
XX
PT   Colon tumor associated proteins and nucleic acids useful for the
PT   prevention, diagnosis and treatment of colonic cancer -
XX
PS   Claim 25; Page 184; 472pp; English.
XX
CC   The present invention describes colon tumour associated proteins (I) and
CC   the polynucleotides (II) that encode them. (I) have cytostatic activity.
CC   (I) and (II) can be used in gene therapy and vaccine production. (I) and
CC   (II) may be used in the prevention, diagnosis and treatment of diseases
CC   associated with inappropriate colon tumour associated protein (TCAP)
CC   expression, such as colonic cancer. For example, (I) and (II) may be
CC   used to treat disorders associated with decreased expression by
CC   rectifying mutations or deletions in a patient's genome that affect the
CC   activity of TCAPs by expressing inactive proteins or to supplement the
CC   patients own production of them. Additionally, (II) may be used to
CC   produce the TCAP proteins, by inserting the nucleic acids into a host
```

```
CC   cell culturing the cell to express the protein. (II) and its
CC   complementary sequences may also be used as DNA probes in diagnostic
CC   polymerase chain reaction (PCR) and hybridisation assays to detect and
CC   quantitate the presence of similar nucleic acids in samples, and
CC   therefore which patients may be in need of restorative therapy. (I) may
CC   also be used as antigens in the production of antibodies against TCAPs
CC   and in assays to identify modulators of TCAP expression and activity.
CC   Anti-(I) antibodies and antagonists may also be used to down regulate
CC   TCAP expression and activity. The anti-(I) antibodies may also be used
CC   as diagnostic agents for detecting the presence of TCAPs in samples
CC   (e.g. by enzyme linked immunosorbant assay (ELISA)). AAI28460 to AAI29512
CC   and AAM24494 to AAM24523 represent nucleotide and amino acid sequences
CC   given in the exemplification of the present invention.
XX
SQ   Sequence 409 BP; 131 A; 104 C; 80 G; 94 T; 0 other;


alignment_scores:
            Quality:  133.00              Length:    133
              Ratio:  1.000                 Gaps:      0
 Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x AAI28641   ..

 Align seg 1/1  to: AAI28641  from: 1  to: 409

      48 TyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLysAspTh 64
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      11 TACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGACAC 60

      64 rLysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTrpGlyA 81
         |||||||||||||||||||||||||||||||||||||||||||||||||
      61 AAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGGGTG 110

      81 spGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerLys 97
         |||||||||||||||||||||||||||||||||||||||||||||||||
     111 ACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAATCCAAG 160

      98 ThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCysProHi 114
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     161 ACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGCCCACA 210

     114 sSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGlnLysL 131
         |||||||||||||||||||||||||||||||||||||||||||||||||
     211 CAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCAGAAAT 260

     131 euAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrAspLys 147
         |||||||||||||||||||||||||||||||||||||||||||||||||
     261 TGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTGACAAA 310

     148 HisLeuSerProAspGlyGlnTyrValProArgIleMetPheValAspPr 164
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     311 CACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTTGACCC 360

     164 oSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnArgLeu 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     361 ATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAATCGTCTC 409
```

```
OM of: US-09-674-266A-181 to: EST:*   out_format : pfs

Date: Oct 8, 2002 11:09 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=x1h
 -Q=/cgn2_1/USPTO_spool/US09674266/runat_08102002_093408_26991/app_query.fasta_1.271
 -DB=EST -QFMT=fastap -SUFFIX=olip2n.rst -GAPOP=4.500
 -GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=quality
 -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext
 -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US09674266_@CGN1_1_2564 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-674-266A-181
Query length: 206
Database: EST:*
Database sequences: 13736207
Database length: -1841457050
Search time (sec): 1725.590000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000


score_list:
Sequence            Strd Orig    ZScore    EScore Len   ! Documentation  ..
gb_est2:BG778248      +  195.00 3425.75  1.4e-181  666   ! BG778248 602666830F1 NIH_MGC_60
gb_est1:AW582256      +  195.00 3425.48  1.5e-181  689   ! AW582256 QV4-ST0212-120100-075-
gb_est2:BI523582      +  194.00 3408.61  1.3e-180  624   ! BI523582 603051712F1 NIH_MGC_12
gb_est2:BG179160      +  191.00 3354.19  1.4e-177  750   ! BG179160 602330032F1 NIH_MGC_91
gb_est2:BE748141      +  190.00 3337.21  1.2e-176  689   ! BE748141 601571509F1 NIH_MGC_55
gb_est2:BM129777      +  189.00 3320.91  9.9e-176  581   ! BM129777 if23a08.y1 Melton Norm
gb_est1:AW956284      +  189.00 3319.20  1.2e-175  721   ! AW956284 EST368354 MAGE reseque
gb_est2:BE796856      +  187.00 3281.72  1.5e-173  948   ! BE796856 601588166F1 NIH_MGC_7
gb_est2:BE870718      +  186.00 3266.09  1.1e-172  734   ! BE870718 601448570F1 NIH_MGC_65
gb_est2:BG285246      +  184.00 3229.47  1.2e-170  866   ! BG285246 602409569F1 NIH_MGC_91
gb_est2:BG386151      +  178.00 3124.68  8.5e-165  751   ! BG386151 602455333F1 NIH_MGC_15
gb_est2:BM006356      +  178.00 3124.61  8.5e-165  758   ! BM006356 603615055F1 NIH_MGC_11
gb_est2:BM006510      +  178.00 3124.21  9.0e-165  797   ! BM006510 603615461F1 NIH_MGC_11
gb_est2:BM006784      +  178.00 3124.16  9.0e-165  802   ! BM006784 603615217F1 NIH_MGC_11
gb_est2:BM006554      +  178.00 3123.74  9.5e-165  846   ! BM006554 603615521F1 NIH_MGC_11
gb_est2:BG610804      +  178.00 3123.67  9.6e-165  853   ! BG610804 602612442F1 NIH_MGC_60
gb_est2:BM006458      +  178.00 3123.58  9.7e-165  863   ! BM006458 603615186F1 NIH_MGC_11
gb_est2:BM006669      +  178.00 3123.34  1.0e-164  890   ! BM006669 603615373F1 NIH_MGC_11
gb_est2:BI763228      +  176.00 3091.12  6.3e-163  603   ! BI763228 603049976F1 NIH_MGC_11
gb_est2:BE867131      +  173.00 3037.90  5.8e-160  623   ! BE867131 601442973F1 NIH_MGC_65
gb_est2:BG823844      +  165.00 2895.28  5.1e-152  742   ! BG823844 602729503F1 NIH_MGC_15
gb_est2:BI770077      +  164.00 2877.15  5.2e-151  788   ! BI770077 603053412F1 NIH_MGC_12
gb_est1:AL543472      +  163.00 2863.22  3.1e-150  493   ! AL543472 AL543472 LTI_NFL006_PI
gb_est2:BI770277      +  162.00 2841.91  4.8e-149  781   ! BI770277 603055968F1 NIH_MGC_12
gb_est1:AW369331      +  161.00 2826.00  3.7e-148  627   ! AW369331 QV1-BN0002-061299-039-
gb_est1:AA314225      +  155.00 2721.31  2.5e-142  537   ! AA314225 EST186091 Colon carcin
gb_est1:AA437001      +  149.00 2616.81  1.6e-136  449   ! AA437001 zv58e07.r1 Soares_test
gb_est1:AA314372      +  149.00 2616.79  1.6e-136  450   ! AA314372 EST186210 Colon carcin
gb_est2:BF207999      +  149.00 2611.48  3.3e-136  879   ! BF207999 601862323F1 NIH_MGC_53
gb_est1:AA314146      +  146.00 2564.04  1.4e-133  438   ! AA314146 EST186014 Colon carcin
gb_est1:AA315629      +  146.00 2563.53  1.5e-133  467   ! AA315629 EST187394 Colon carcin
gb_est1:AW364300      -  146.00 2561.07  2.1e-133  637   ! AW364300 QV3-DT0012-291299-051-
gb_est2:BG259927      +  142.00 2487.80  2.5e-129  890   ! BG259927 602371889F1 NIH_MGC_93
gb_est2:BF747277      +  140.00 2458.27  1.1e-127  430   ! BF747277 RC1-BT0254-271100-121-
gb_est2:BF513917      -  140.00 2457.25  1.3e-127  489   ! BF513917 UI-H-BW1-amq-f-02-0-UI
gb_est1:AA316233      +  138.00 2422.78  1.1e-125  440   ! AA316233 EST187946 HCC cell lin
gb_est2:BF733142      +  137.00 2400.84  1.8e-124  755   ! BF733142 EST039 Human hepatocel
gb_est1:AA315724      +  133.00 2334.49  8.7e-121  441   ! AA315724 EST187482 Colon carcin
gb_est1:AA316967      +  133.00 2334.47  8.8e-121  442   ! AA316967 EST188638 HCC cell
gb_est1:AA316115      +  133.00 2332.51  1.1e-120  566   ! AA316115 EST187838 Colon car
gb_est1:AA315166      +  128.00 2246.70  6.8e-116  415   ! AA315166 EST186963 HCC cell
gb_est1:AW166169      -  124.00 2172.91  8.7e-112  619   ! AW166169 xf46b06.x1 NCI_CGAP
gb_est2:BG501149      +  123.00 2153.84  1.0e-110  740   ! BG501149 602546389F1 NIH_MGC
gb_est2:BF680021      +  123.00 2153.77  1.0e-110  747   ! BF680021 602154740F1 NIH_MGC
gb_est2:BF680376      +  123.00 2152.72  1.2e-110  853   ! BF680376 602154161F1 NIH_MGC

seq_name: gb_est2:BG778248

seq_documentation_block:
LOCUS       BG778248                666 bp    mRNA    linear   EST 15-MAY-2001
DEFINITION  602666830F1 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:4806679 5',
            mRNA sequence.
ACCESSION   BG778248
VERSION     BG778248.1  GI:14048565
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 666)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1656  row: 1  column: 08
            High quality sequence stop: 665.
FEATURES             Location/Qualifiers
     source          1. .666
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4806679"
                     /clone_lib="NIH_MGC_60"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: prostate; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                     ); Double-stranded cDNA was prepared from cell line RNA.
                     5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).  Average
                     insert size 1.5 kb (range 0.9-4.0 kb). 14/15 colonies
                     contained inserts  by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
                     Library."
BASE COUNT      205 a    165 c    140 g    156 t
ORIGIN


alignment_scores:
            Quality:  195.00              Length:    195
              Ratio:  1.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x BG778248   ..

 Align seg 1/1  to: BG778248  from: 1  to: 666

      12 GlyProHisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSe 28
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      12 GGACCGCATCCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATC 61
```

```
     28 rArgValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValA 45
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     62 CAGAGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGG 111

     45 laLeuSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLys 61
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    112 CCCTCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAA 161

     62 LysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGl 78
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    162 AAGGACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGG 211

     78 yTrpGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrL 95
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    212 TTGGGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATA 261

     95 ysSerLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGlu 111
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    262 AATCCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAG 311

    112 CysProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIl 128
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    312 TGCCCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAAT 361

    128 eGlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrT 145
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    362 CCAGAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAA 411

    145 hrAspLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPhe 161
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    412 CTGACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTT 461

    162 ValAspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAs 178
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    462 GTTGACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAA 511

    178 nArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnM 195
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    512 TCGTCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACA 561

    195 etLysLysAlaLeuLysLeuLeuLysThrGluLeu 206
        |||||||||||||||||||||||||||||||||||
    562 TGAAGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 596

seq_name: gb_est1:AW582256

seq_documentation_block:
LOCUS       AW582256               689 bp    mRNA    linear   EST 16-MAR-2000
DEFINITION  QV4-ST0212-120100-075-e10 ST0212 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW582256
VERSION     AW582256.1  GI:7257305
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 689)
  AUTHORS   HCGP http://www.ludwig.org.br/ORESTES.
  TITLE     The FAPESP/LICR Human Cancer Genome Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV4&t2=QV4-ST0212-
            120100-075-e10&t3=2000-01-12&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 687.
FEATURES             Location/Qualifiers
     source          1. .689
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="ST0212"
                     /dev_stage="Adult"
                     /note="Organ: stomach; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      212 a    172 c    143 g    162 t
ORIGIN


alignment_scores:
            Quality:  195.00             Length:    195
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x AW582256   ..

 Align seg 1/1  to: AW582256  from: 1  to: 689

     12 GlyProHisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSe 28
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     25 GGACCGCATCCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATC 74

     28 rArgValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValA 45
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     75 CAGAGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGG 124

     45 laLeuSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLys 61
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    125 CCCTCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAA 174

     62 LysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGl 78
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    175 AAGGACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGG 224

     78 yTrpGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrL 95
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    225 TTGGGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATA 274

     95 ysSerLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGlu 111
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    275 AATCCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAG 324

    112 CysProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIl 128
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    325 TGCCCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAAT 374

    128 eGlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrT 145
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    375 CCAGAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAA 424

    145 hrAspLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPhe 161
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    425 CTGACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTT 474

    162 ValAspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAs 178
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    475 GTTGACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAA 524

    178 nArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnM 195
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    525 TCGTCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACA 574
```

```
     195 etLysLysAlaLeuLysLeuLeuLysThrGluLeu 206
         ||||||||||||||||||||||||||||||||||
     575 TGAAGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 609

seq_name: gb_est2:BI523582

seq_documentation_block:
LOCUS       BI523582               624 bp    mRNA    linear   EST 29-AUG-2001
DEFINITION  603051712F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5201331 5',
            mRNA sequence.
ACCESSION   BI523582
VERSION     BI523582.1  GI:15348374
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 624)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11504  row: h  column: 04
            High quality sequence start: 5
            High quality sequence stop: 624.
FEATURES             Location/Qualifiers
     source          1. .624
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5201331"
                     /clone_lib="NIH_MGC_122"
                     /lab_host="DH10B"
                     /note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
                     Site_1: NotI; Site_2: EcoRV (destroyed); RNA source
                     anonymous pool of 24 week female lung, 16 week female
                     spleen, and 20-22 week male spleens.  Library is oligo-dT
                     primed and directionally cloned (EcoRV site is destroyed
                     upon  cloning).  Average insert size 1.4 kb, insert size
                     range 1-3 kb. Library is normalized and enriched for
                     full-length clones and was  constructed by C. Gruber
                     (Invitrogen).  Research Genetics tracking  code 026. Note:
                     this is a NIH_MGC Library."
BASE COUNT      190 a    161 c    126 g    147 t
ORIGIN


alignment_scores:
           Quality:  194.00            Length:    194
             Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x BI523582   ..

 Align seg 1/1  to: BI523582  from: 1  to: 624

      13 ProHisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerAr 29
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       2 CCGCATCCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAG 51

      29 gValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAlaL 46
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      52 AGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCC 101

      46 euSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLys 62
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     102 TCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAG 151

      63 AspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTr 79
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     152 GACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTG 201

      79 pGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysS 96
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     202 GGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAAT 251

      96 erLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCys 112
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     252 CCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGC 301

     113 ProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGl 129
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     302 CCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCA 351

     129 nLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrA 146
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     352 GAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTG 401

     146 spLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheVal 162
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     402 ACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTT 451

     163 AspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnAr 179
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     452 GACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAATCG 501

     179 gLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetL 196
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     502 TCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACATGA 551

     196 ysLysAlaLeuLysLeuLeuLysThrGluLeu 206
         |||||||||||||||||||||||||||||||
     552 AGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 583

seq_name: gb_est2:BG179160

seq_documentation_block:
LOCUS       BG179160               750 bp    mRNA    linear   EST 06-FEB-2001
DEFINITION  602330032F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4431583 5',
            mRNA sequence.
ACCESSION   BG179160
VERSION     BG179160.1  GI:12685943
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 750)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10186  row: o  column: 08
            High quality sequence stop: 732.
FEATURES             Location/Qualifiers
     source          1. .750
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4431583"
                     /clone_lib="NIH_MGC_91"
```

```
                     /tissue_type="adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: prostate; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.4 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      234 a    174 c    158 g    183 t      1 others
ORIGIN


alignment_scores:
            Quality:  191.00             Length:    191
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x BG179160   ..

 Align seg 1/1  to: BG179160  from: 1  to: 750

      15 ProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerArgValAl 31
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      15 CCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAGAGTTGC 64

      31 aMetGluLysIleProValSerAlaPheLeuLeuLeuValAlaLeuSerT 48
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      65 CATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCTCCT 114

      48 yrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLysAspThr 64
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     115 ACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGACACA 164

      65 LysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTrpGlyAs 81
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     165 AAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGGGTGA 214

      81 pGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerLysT 98
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     215 CCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAATCCAAGA 264

      98 hrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCysProHis 114
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     265 CAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGCCCACAC 314

     115 SerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGlnLysLe 131
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     315 AGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCAGAAATT 364

     131 uAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrAspLysH 148
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     365 GGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTGACAAAC 414

     148 isLeuSerProAspGlyGlnTyrValProArgIleMetPheValAspPro 164
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     415 ACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTTGACCCA 464

     165 SerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnArgLeuTy 181
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     465 TCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAATCGTCTCTA 514

     181 rAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLysA 198
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     515 TGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACATGAAGAAAG 564

     198 laLeuLysLeuLeuLysThrGlu 205
         |||||||||||||||||||||||
     565 CTCTCAAGTTGCTGAAGACTGAA 587

seq_name: gb_est2:BE748141

seq_documentation_block:
LOCUS       BE748141                689 bp    mRNA    linear   EST 15-SEP-2000

DEFINITION  601571509F1 NIH_MGC_55 Homo sapiens cDNA clone IMAGE:3838347 5',
            mRNA sequence.
ACCESSION   BE748141
VERSION     BE748141.1  GI:10162133
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 689)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM525  row: a  column: 04
            High quality sequence stop: 643.
FEATURES             Location/Qualifiers
     source          1. .689
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3838347"
                     /clone_lib="NIH_MGC_55"
                     /tissue_type="from acute myelogenous leukemia"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: bone marrow; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                     ); Double-stranded cDNA was prepared from cell line RNA.
                     5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).   Average
                     insert size 1.65 kb (range 0.9-4.0 kb).  14/15 colonies
                     contained  inserts by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT      209 a    170 c    148 g    162 t
ORIGIN


alignment_scores:
            Quality:  190.00             Length:    190
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x BE748141   ..

 Align seg 1/1  to: BE748141  from: 1  to: 689

      15 ProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerArgValAl 31
         |||||||||||||||||||||||||||||||||||||||||||||||||||
       3 CCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAGAGTTGC 52

      31 aMetGluLysIleProValSerAlaPheLeuLeuLeuValAlaLeuSerT 48
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      53 CATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCTCCT 102

      48 yrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLysAspThr 64
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     103 ACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGACACA 152

      65 LysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTrpGlyAs 81
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     153 AAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGGGTGA 202
```

```
     81 pGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerLysT 98
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    203 CCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAATCCAAGA 252

     98 hrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCysProHis 114
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    253 CAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGCCCACAC 302

    115 SerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGlnLysLe 131
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    303 AGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCAGAAATT 352

    131 uAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrAspLysH 148
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    353 GGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTGACAAAC 402

    148 isLeuSerProAspGlyGlnTyrValProArgIleMetPheValAspPro 164
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    403 ACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTTGACCCA 452

    165 SerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnArgLeuTy 181
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    453 TCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAACCGTCTCTA 502

    181 rAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLysA 198
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    503 TGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACATGAAGAAAG 552

    198 laLeuLysLeuLeuLysThr 204
        ||||||||||||||||||
    553 CTCTCAAGTTGCTGAAGACT 572
```

seq_name: gb_est2:BM129777

seq_documentation_block:

```
LOCUS       BM129777                581 bp    mRNA    linear   EST 27-NOV-2001
DEFINITION  if23a08.y1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
            cDNA 5' similar to TR:O95994 O95994 SECRETED CEMENT GLAND PROTEIN
            XAG-2 HOMOLOG. ;, mRNA sequence.
ACCESSION   BM129777
VERSION     BM129777.1  GI:17124329
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 581)
  AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
            ,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.
            , Jackson,Y. and Bowers,Y.
  TITLE     Endocrine Pancreas Consortium
  JOURNAL   Unpublished (2000)
COMMENT     Other_ESTs: if23a08.x1
            Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Douglas Melton DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Juliana Brown
            (brown@fas.harvard.edu)
            High quality sequence stop: 429.
FEATURES             Location/Qualifiers
     source          1. .581
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
                     /sex="Both"
                     /tissue_type="Islets of Langerhans"
                     /dev_stage="Adult"
                     /lab_host="DH10B"
                     /note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;
                     Site_2: Sal 1; Starting library constructed using
                     SuperScript Plasmid Library kit (Life Technologies). cDNA
                     made by oligo-dT priming. Size-selected by column
                     fractionation; average insert size 1.08 kb. Library was
                     amplified once on solid support and plasmid DNA from
                     library was prepared. The library DNA was normalized by
                     method #4 from Bonaldo, Lennon, and Soares 1996 Genome
                     Research 6:791-806; 0.5 microgram single-stranded library
                     plasmid DNA was mixed with 5 micrograms PCR product
                     representing library inserts and hybridized to an Ecot of
                     20. Single-stranded (unhybridized) plasmids were isolated
                     by hydroxyapatite chromatography and used to make this
                     library."
BASE COUNT      185 a    140 c    118 g    138 t
ORIGIN
```

alignment_scores:

```
            Quality:  189.00              Length:    189
              Ratio:  1.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000
```

alignment_block:

US-09-674-266A-181 x BM129777 ..

Align seg 1/1 to: BM129777 from: 1 to: 581

```
     18 ArgLeuThrGlnGlyArgTrpValArgLysSerArgValAlaMetGluLy 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 CGACTCACACAAGGCAGGTGGGTGAGGAAATCCAGAGTTGCCATGGAGAA 50

     34 sIleProValSerAlaPheLeuLeuLeuValAlaLeuSerTyrThrLeuA 51
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     51 AATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCTCCTACACTCTGG 100

     51 laArgAspThrThrValLysProGlyAlaLysLysAspThrLysAspSer 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    101 CCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGACACAAAGGACTCT 150

     68 ArgProLysLeuProGlnThrLeuSerArgGlyTrpGlyAspGlnLeuIl 84
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    151 CGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGGGTGACCAACTCAT 200

     84 eTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerLysThrSerAsnL 101
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    201 CTGGACTCAGACATATGAAGAAGCTCTATATAAATCCAAGACAAGCAACA 250

    101 ysProLeuMetIleIleHisHisLeuAspGluCysProHisSerGlnAla 117
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    251 AACCCTTGATGATTATTCATCACTTGGATGAGTGCCCACACAGTCAAGCT 300

    118 LeuLysLysValPheAlaGluAsnLysGluIleGlnLysLeuAlaGluGl 134
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    301 TTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCAGAAATTGGCAGAGCA 350

    134 nPheValLeuLeuAsnLeuValTyrGluThrThrAspLysHisLeuSerP 151
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    351 GTTTGTCCTCCTCAATCTGGTTTATGAAACAACTGACAAACACCTTTCTC 400

    151 roAspGlyGlnTyrValProArgIleMetPheValAspProSerLeuThr 167
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    401 CTGATGGCCAGTATGTCCCCAGGATTATGTTTGTTGACCCATCTCTGACA 450

    168 ValArgAlaAspIleThrGlyArgTyrSerAsnArgLeuTyrAlaTyrGl 184
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    451 GTTAGAGCCGATATCACTGGAAGATATTCAAATCGTCTCTATGCTTACGA 500
```

```
     184 uProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLysAlaLeuLysL 201
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     501 ACCTGCAGATACAGCTCTGTTGCTTGACAACATGAAGAAAGCTCTCAAGT 550

     201 euLeuLysThrGluLeu 206
         ||||||||||||||||
     551 TGCTGAAGACTGAATTG 567

seq_name: gb_est1:AW956284

seq_documentation_block:
LOCUS       AW956284                721 bp    mRNA    linear   EST 01-JUN-2000
DEFINITION  EST368354 MAGE resequences, MAGD Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW956284
VERSION     AW956284.1  GI:8145967
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 721)
  AUTHORS   Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
            ,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
            Quackenbush,J.
  TITLE     Assessment of gene expression patterns in a model of colon tumor
            metastasis using a 19,200 element cDNA microarray
  JOURNAL   Unpublished (2000)
COMMENT     Contact: John Quackenbush
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 3528
            Fax: 301 838 0208
            Email: johnq@tigr.org
            Plate: 90
            Seq primer: Reverse.
FEATURES             Location/Qualifiers
     source          1. .721
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="MAGE resequences, MAGD"
                     /note="Vector: pBluescriptSKm"
BASE COUNT      224 a    173 c    148 g    175 t      1 others
ORIGIN


alignment_scores:
            Quality:  189.00             Length:    189
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x AW956284   ..

 Align seg 1/1  to: AW956284  from: 1  to: 721

      18 ArgLeuThrGlnGlyArgTrpValArgLysSerArgValAlaMetGluLy 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      21 CGACTCACACAAGGCAGGTGGGTGAGGAAATCCAGAGTTGCCATGGAGAA 70

      34 sIleProValSerAlaPheLeuLeuLeuValAlaLeuSerTyrThrLeuA 51
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      71 AATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCTCCTACACTCTGG 120

      51 laArgAspThrThrValLysProGlyAlaLysLysAspThrLysAspSer 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     121 CCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGACACAAAGGACTCT 170

      68 ArgProLysLeuProGlnThrLeuSerArgGlyTrpGlyAspGlnLeuIl 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     171 CGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGGGTGACCAACTCAT 220

      84 eTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerLysThrSerAsnL 101
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     221 CTGGACTCAGACATATGAAGAAGCTCTATATAAATCCAAGACAAGCAACA 270

     101 ysProLeuMetIleIleHisHisLeuAspGluCysProHisSerGlnAla 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     271 AACCCTTGATGATTATTCATCACTTGGATGAGTGCCCACACAGTCAAGCT 320

     118 LeuLysLysValPheAlaGluAsnLysGluIleGlnLysLeuAlaGluGl 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     321 TTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCAGAAATTGGCAGAGCA 370

     134 nPheValLeuLeuAsnLeuValTyrGluThrThrAspLysHisLeuSerP 151
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     371 GTTTGTCCTCCTCAATCTGGTTTATGAAACAACTGACAAACACCTTTCTC 420

     151 roAspGlyGlnTyrValProArgIleMetPheValAspProSerLeuThr 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     421 CTGATGGCCAGTATGTCCCCAGGATTATGTTTGTTGACCCATCTCTGACA 470

     168 ValArgAlaAspIleThrGlyArgTyrSerAsnArgLeuTyrAlaTyrGl 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     471 GTTAGAGCCGATATCACTGGAAGATATTCAAATCGTCTCTATGCTTACGA 520

     184 uProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLysAlaLeuLysL 201
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     521 ACCTGCAGATACAGCTCTGTTGCTTGACAACATGAAGAAAGCTCTCAAGT 570

     201 euLeuLysThrGluLeu 206
         ||||||||||||||||
     571 TGCTGAAGACTGAATTG 587

seq_name: gb_est2:BE796856

seq_documentation_block:
LOCUS       BE796856                948 bp    mRNA    linear   EST 20-SEP-2000
DEFINITION  601588166F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3942112 5',
            mRNA sequence.
ACCESSION   BE796856
VERSION     BE796856.1  GI:10218156
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 948)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM795  row: d  column: 17
            High quality sequence stop: 748.
FEATURES             Location/Qualifiers
     source          1. .948
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3942112"
                     /clone_lib="NIH_MGC_7"
                     /tissue_type="small cell carcinoma"
                     /cell_line="MGC3"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
```

```
                    (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      296 a    225 c    207 g    220 t
ORIGIN


alignment_scores:
            Quality:  187.00             Length:    187
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x BE796856   ..

 Align seg 1/1  to: BE796856  from: 1  to: 948

      20 ThrGlnGlyArgTrpValArgLysSerArgValAlaMetGluLysIlePr 36
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       3 ACACAAGGCAGGTGGGTGAGGAAATCCAGAGTTGCCATGGAGAAAATTCC 52

      36 oValSerAlaPheLeuLeuLeuValAlaLeuSerTyrThrLeuAlaArgA 53
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      53 AGTGTCAGCATTCTTGCTCCTTGTGGCCCTCTCCTACACTCTGGCCAGAG 102

      53 spThrThrValLysProGlyAlaLysLysAspThrLysAspSerArgPro 69
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     103 ATACCACAGTCAAACCTGGAGCCAAAAAGGACACAAAGGACTCTCGACCC 152

      70 LysLeuProGlnThrLeuSerArgGlyTrpGlyAspGlnLeuIleTrpTh 86
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     153 AAACTGCCCCAGACCCTCTCCAGAGGTTGGGGTGACCAACTCATCTGGAC 202

      86 rGlnThrTyrGluGluAlaLeuTyrLysSerLysThrSerAsnLysProL 103
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     203 TCAGACATATGAAGAAGCTCTATATAAATCCAAGACAAGCAACAAACCCT 252

     103 euMetIleIleHisHisLeuAspGluCysProHisSerGlnAlaLeuLys 119
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     253 TGATGATTATTCATCACTTGGATGAGTGCCCACACAGTCAAGCTTTAAAG 302

     120 LysValPheAlaGluAsnLysGluIleGlnLysLeuAlaGluGlnPheVa 136
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     303 AAAGTGTTTGCTGAAAATAAAGAAATCCAGAAATTGGCAGAGCAGTTTGT 352

     136 lLeuLeuAsnLeuValTyrGluThrThrAspLysHisLeuSerProAspG 153
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     353 CCTCCTCAATCTGGTTTATGAAACAACTGACAAACACCTTTCTCCTGATG 402

     153 lyGlnTyrValProArgIleMetPheValAspProSerLeuThrValArg 169
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     403 GCCAGTATGTCCCCAGGATTATGTTTGTTGACCCATCTCTGACAGTTAGA 452

     170 AlaAspIleThrGlyArgTyrSerAsnArgLeuTyrAlaTyrGluProAl 186
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     453 GCCGATATCACTGGAAGATATTCAAACCGTCTCTATGCTTACGAACCTGC 502

     186 aAspThrAlaLeuLeuLeuAspAsnMetLysLysAlaLeuLysLeuLeuL 203
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     503 AGATACAGCTCTGTTGCTTGACAACATGAAGAAAGCTCTCAAGTTGCTGA 552

     203 ysThrGluLeu 206
         ||||||||||
     553 AGACTGAATTG 563

seq_name: gb_est2:BE870718

seq_documentation_block:
LOCUS       BE870718               734 bp    mRNA    linear   EST 20-OCT-2000
DEFINITION  601448570F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3852448 5',
            mRNA sequence.
ACCESSION   BE870718
VERSION     BE870718.1  GI:10319507
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 734)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9574  row: l  column: 17
            High quality sequence stop: 723.
FEATURES             Location/Qualifiers
     source          1. .734
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3852448"
                     /clone_lib="NIH_MGC_65"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.8 kb.   Library constructed by Life
                     Technologies. "
BASE COUNT      231 a    163 c    156 g    184 t
ORIGIN


alignment_scores:
            Quality:  186.00             Length:    186
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x BE870718   ..

 Align seg 1/1  to: BE870718  from: 1  to: 734

      21 GlnGlyArgTrpValArgLysSerArgValAlaMetGluLysIleProVa 37
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       2 CAAGGCAGGTGGGTGAGGAAATCCAGAGTTGCCATGGAGAAAATTCCAGT 51

      37 lSerAlaPheLeuLeuLeuValAlaLeuSerTyrThrLeuAlaArgAspT 54
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      52 GTCAGCATTCTTGCTCCTTGTGGCCCTCTCCTACACTCTGGCCAGAGATA 101

      54 hrThrValLysProGlyAlaLysLysAspThrLysAspSerArgProLys 70
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     102 CCACAGTCAAACCTGGAGCCAAAAAGGACACAAAGGACTCTCGACCCAAA 151

      71 LeuProGlnThrLeuSerArgGlyTrpGlyAspGlnLeuIleTrpThrGl 87
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     152 CTGCCCCAGACCCTCTCCAGAGGTTGGGGTGACCAACTCATCTGGACTCA 201

      87 nThrTyrGluGluAlaLeuTyrLysSerLysThrSerAsnLysProLeuM 104
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     202 GACATATGAAGAAGCTCTATATAAATCCAAGACAAGCAACAAACCCTTGA 251

     104 etIleIleHisHisLeuAspGluCysProHisSerGlnAlaLeuLysLys 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     252 TGATTATTCATCACTTGGATGAGTGCCCACACAGTCAAGCTTTAAAGAAA 301

     121 ValPheAlaGluAsnLysGluIleGlnLysLeuAlaGluGlnPheValLe 137
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     302 GTGTTTGCTGAAAATAAAGAAATCCAGAAATTGGCAGAGCAGTTTGTCCT 351

     137 uLeuAsnLeuValTyrGluThrThrAspLysHisLeuSerProAspGlyG 154
```

```
     ||||||||||||||||||||||||||||||||||||||||||||||||
 352 CCTCAATCTGGTTTATGAAACAACTGACAAACACCTTTCTCCTGATGGCC 401

 154 lnTyrValProArgIleMetPheValAspProSerLeuThrValArgAla 170
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 402 AGTATGTCCCCAGGATTATGTTTGTTGACCCATCTCTGACAGTTAGAGCC 451

 171 AspIleThrGlyArgTyrSerAsnArgLeuTyrAlaTyrGluProAlaAs 187
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 452 GATATCACTGGAAGATATTCAAACCGTCTCTATGCTTACGAACCTGCAGA 501

 187 pThrAlaLeuLeuLeuAspAsnMetLysLysAlaLeuLysLeuLeuLysT 204
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 502 TACAGCTCTGTTGCTTGACAACATGAAGAAAGCTCTCAAGTTGCTGAAGA 551

 204 hrGluLeu 206
     ||||||||
 552 CTGAATTG 559

seq_name: gb_est2:BG285246

seq_documentation_block:
LOCUS       BG285246                866 bp    mRNA    linear   EST 21-FEB-2001
DEFINITION  602409569F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4539353 5',
            mRNA sequence.
ACCESSION   BG285246
VERSION     BG285246.1  GI:13037011
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 866)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10467  row: i  column: 18
            High quality sequence stop: 750.
FEATURES             Location/Qualifiers
     source          1. .866
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4539353"
                     /clone_lib="NIH_MGC_91"
                     /tissue_type="adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: prostate; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.4 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      270 a    194 c    189 g    212 t      1 others
ORIGIN


alignment_scores:
            Quality:  184.00             Length:    184
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x BG285246   ..

 Align seg 1/1  to: BG285246  from: 1  to: 866
```

```
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  17 ArgArgLeuThrGlnGlyArgTrpValArgLysSerArgValAlaMetGl 33
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  30 CGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAGAGTTGCCATGGA 79

  33 uLysIleProValSerAlaPheLeuLeuLeuValAlaLeuSerTyrThrL 50
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  80 GAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCTCCTACACTC 129

  50 euAlaArgAspThrThrValLysProGlyAlaLysLysAspThrLysAsp 66
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 130 TGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGACACAAAGGAC 179

  67 SerArgProLysLeuProGlnThrLeuSerArgGlyTrpGlyAspGlnLe 83
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 180 TCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGGGTGACCAACT 229

  83 uIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerLysThrSerA 100
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 230 CATCTGGACTCAGACATATGAAGAAGCTCTATATAAATCCAAGACAAGCA 279

 100 snLysProLeuMetIleIleHisHisLeuAspGluCysProHisSerGln 116
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 280 ACAAACCCTTGATGATTATTCATCACTTGGATGAGTGCCCACACAGTCAA 329

 117 AlaLeuLysLysValPheAlaGluAsnLysGluIleGlnLysLeuAlaGl 133
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 330 GCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCAGAAATTGGCAGA 379

 133 uGlnPheValLeuLeuAsnLeuValTyrGluThrThrAspLysHisLeuS 150
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 380 GCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTGACAAACACCTTT 429

 150 erProAspGlyGlnTyrValProArgIleMetPheValAspProSerLeu 166
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 430 CTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTTGACCCATCTCTG 479

 167 ThrValArgAlaAspIleThrGlyArgTyrSerAsnArgLeuTyrAlaTy 183
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 480 ACAGTTAGAGCCGATATCACTGGAAGATATTCAAATCGTCTCTATGCTTA 529

 183 rGluProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLysAlaLeuL 200
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 530 CGAACCTGCAGATACAGCTCTGTTGCTTGACAACATGAAGAAAGCTCTCA 579

 200 ys 200
     ||
 580 AG 581

seq_name: gb_est2:BG386151

seq_documentation_block:
LOCUS       BG386151                751 bp    mRNA    linear   EST 12-MAR-2001
DEFINITION  602455333F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4583336 5',
            mRNA sequence.
ACCESSION   BG386151
VERSION     BG386151.1  GI:13279597
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 751)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: NIH Intramural Sequencing Center
             Clone distribution: MGC clone distribution information can be
```

```
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLCM1307  row: b  column: 09
          High quality sequence stop: 672.
FEATURES             Location/Qualifiers
     source          1. .751
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4583336"
                     /clone_lib="NIH_MGC_15"
                     /tissue_type="adenocarcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT      230 a    181 c    153 g    187 t
ORIGIN


alignment_scores:
            Quality:  178.00             Length:    178
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x BG386151   ..

 Align seg 1/1  to: BG386151  from: 1  to: 751

      29 ArgValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAl 45
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      26 AGAGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGC 75

      45 aLeuSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysL 62
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      76 CCTCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAA 125

      62 ysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGly 78
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     126 AGGACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGT 175

      79 TrpGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLy 95
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     176 TGGGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAA 225

      95 sSerLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluC 112
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     226 ATCCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGT 275

     112 ysProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIle 128
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     276 GCCCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATC 325

     129 GlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrTh 145
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     326 CAGAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAAC 375

     145 rAspLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheV 162
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     376 TGACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTG 425

     162 alAspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsn 178
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     426 TTGACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAAC 475

     179 ArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMe 195
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     476 CGTCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACAT 525

     195 tLysLysAlaLeuLysLeuLeuLysThrGluLeu 206
         |||||||||||||||||||||||||||||||||
     526 GAAGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 559

seq_name: gb_est2:BM006356

seq_documentation_block:
LOCUS       BM006356                758 bp    mRNA    linear   EST 30-OCT-2001
DEFINITION  603615055F1 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:5421109 5',
            mRNA sequence.
ACCESSION   BM006356
VERSION     BM006356.1  GI:16520710
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 758)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1876  row: m  column: 14
            High quality sequence stop: 754.
FEATURES             Location/Qualifiers
     source          1. .758
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5421109"
                     /clone_lib="NIH_MGC_110"
                     /tissue_type="ductal carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Library constructed  by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of  California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      238 a    176 c    156 g    188 t
ORIGIN


alignment_scores:
            Quality:  178.00             Length:    178
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x BM006356   ..

 Align seg 1/1  to: BM006356  from: 1  to: 758

      29 ArgValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAl 45
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      31 AGAGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGC 80

      45 aLeuSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysL 62
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      81 CCTCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAA 130

      62 ysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGly 78
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
     131 AGGACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGT 180

      79 TrpGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLy 95
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     181 TGGGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAA 230

      95 sSerLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluC 112
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     231 ATCCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGT 280

     112 ysProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIle 128
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     281 GCCCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATC 330

     129 GlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrTh 145
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     331 CAGAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAAC 380

     145 rAspLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheV 162
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     381 TGACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTG 430

     162 alAspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsn 178
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     431 TTGACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAAC 480

     179 ArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMe 195
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     481 CGTCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACAT 530

     195 tLysLysAlaLeuLysLeuLeuLysThrGluLeu 206
         ||||||||||||||||||||||||||||||||||
     531 GAAGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 564
```

seq_name: gb_est2:BM006510

seq_documentation_block:

```
LOCUS       BM006510                797 bp    mRNA    linear   EST 30-OCT-2001
DEFINITION  603615461F1 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:5421399 5',
            mRNA sequence.
ACCESSION   BM006510
VERSION     BM006510.1  GI:16520864
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 797)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1877  row: i  column: 16
            High quality sequence stop: 792.
FEATURES             Location/Qualifiers
     source          1. .797
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5421399"
                     /clone_lib="NIH_MGC_110"
                     /tissue_type="ductal carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Library constructed  by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of  California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library:"
BASE COUNT      258 a    177 c    159 g    203 t
ORIGIN
```

alignment_scores:

```
            Quality:  178.00             Length:    178
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000
```

alignment_block:

US-09-674-266A-181 x BM006510 ..

Align seg 1/1 to: BM006510 from: 1 to: 797

```
      29 ArgValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAl 45
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      31 AGAGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGC 80

      45 aLeuSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysL 62
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      81 CCTCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAA 130

      62 ysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGly 78
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     131 AGGACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGT 180

      79 TrpGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLy 95
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     181 TGGGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAA 230

      95 sSerLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluC 112
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     231 ATCCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGT 280

     112 ysProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIle 128
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     281 GCCCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATC 330

     129 GlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrTh 145
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     331 CAGAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAAC 380

     145 rAspLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheV 162
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     381 TGACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTG 430

     162 alAspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsn 178
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     431 TTGACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAAC 480

     179 ArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMe 195
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     481 CGTCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACAT 530

     195 tLysLysAlaLeuLysLeuLeuLysThrGluLeu 206
         ||||||||||||||||||||||||||||||||||
     531 GAAGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 564
```

seq_name: gb_est2:BM006784

seq_documentation_block:

```
LOCUS       BM006784                802 bp    mRNA    linear   EST 30-OCT-2001
DEFINITION  603615217F1 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:5420836 5',
            mRNA sequence.
ACCESSION   BM006784
VERSION     BM006784.1  GI:16521138
KEYWORDS    EST.
SOURCE      human.
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 802)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1876  row: b  column: 05
            High quality sequence stop: 795.
FEATURES             Location/Qualifiers
     source          1. .802
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5420836"
                     /clone_lib="NIH_MGC_110"
                     /tissue_type="ductal carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Library constructed  by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of  California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      257 a    180 c    162 g    203 t
ORIGIN


alignment_scores:
           Quality:  178.00            Length:    178
             Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x BM006784   ..

 Align seg 1/1  to: BM006784  from: 1  to: 802

      29 ArgValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAl 45
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      31 AGAGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGC 80

      45 aLeuSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysL 62
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      81 CCTCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAA 130

      62 ysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGly 78
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     131 AGGACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGT 180

      79 TrpGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLy 95
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     181 TGGGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAA 230

      95 sSerLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluC 112
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     231 ATCCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGT 280

     112 ysProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIle 128
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     281 GCCCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATC 330

     129 GlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrTh 145
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     331 CAGAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAAC 380

     145 rAspLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheV 162
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     381 TGACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTCG 430

     162 alAspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsn 178
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     431 TTGACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAAC 480

     179 ArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMe 195
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     481 CGTCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACAT 530

     195 tLysLysAlaLeuLysLeuLeuLysThrGluLeu 206
         |||||||||||||||||||||||||||||||||
     531 GAAGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 564

seq_name: gb_est2:BM006554

seq_documentation_block:
LOCUS       BM006554                 846 bp    mRNA    linear   EST 30-OCT-2001
DEFINITION  603615521F1 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:5421413 5',
            mRNA sequence.
ACCESSION   BM006554
VERSION     BM006554.1  GI:16520908
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 846)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1877  row: j  column: 06
            High quality sequence stop: 781.
FEATURES             Location/Qualifiers
     source          1. .846
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5421413"
                     /clone_lib="NIH_MGC_110"
                     /tissue_type="ductal carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Library constructed  by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of  California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      278 a    192 c    173 g    203 t
ORIGIN


alignment_scores:
           Quality:  178.00            Length:    178
             Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
```

US-09-674-266A-181 x BM006554 ..

Align seg 1/1 to: BM006554 from: 1 to: 846

```
     29 ArgValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAl 45
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     31 AGAGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGC 80

     45 aLeuSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysL 62
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     81 CCTCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAA 130

     62 ysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGly 78
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    131 AGGACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGT 180

     79 TrpGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLy 95
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    181 TGGGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAA 230

     95 sSerLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluC 112
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    231 ATCCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGT 280

    112 ysProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIle 128
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    281 GCCCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATC 330

    129 GlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrTh 145
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    331 CAGAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAAC 380

    145 rAspLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheV 162
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    381 TGACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTG 430

    162 alAspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsn 178
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    431 TTGACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAAC 480

    179 ArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMe 195
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    481 CGTCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACAT 530

    195 tLysLysAlaLeuLysLeuLeuLysThrGluLeu 206
        |||||||||||||||||||||||||||||||||
    531 GAAGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 564
```

```
OM of: US-09-674-266A-181 to: Issued_Patents_NA:*   out_format : pfs

Date: Oct 8, 2002 11:10 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlh
 -Q=/cgn2_1/USPTO_spool/US09674266/runat_08102002_093409_27017/app_query.fasta_1.271
 -DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=olip2n.rni
 -GAPOP=4.500 -GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000
 -LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000
 -XGAPEXT=60.000 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000
 -YGAPEXT=60.000 -DELOP=6.000 -DELEXT=7.000 -START=1
 -MATRIX=oligo -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US09674266_@CGN1_1_46 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-674-266A-181
Query length: 206
Database: Issued_Patents_NA:*
Database sequences: 383533
Database length: 122816752
Search time (sec): 46.220000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000


score_list:
Sequence            Strd Orig    ZScore    EScore Len  ! Documentation  ..
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-1 +  192.00 3613.53  1.4e-193  875  !
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-247-155-61 +  178.00 3343.21  1.6e-178 1689  !
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-35 +  107.00 2009.44  3.1e-104  386  !
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-040-984-78 +  102.00 1909.16  1.2e-98   793  !
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-123-912-78 +  102.00 1909.16  1.2e-98   793  !
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-221-298-32 +   66.00 1232.37  5.9e-61   401  !
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-247-155-149 +   62.00 1150.01  2.3e-56   940  !
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-36 +   53.00  986.63  2.9e-47   373  !
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-030-607-209 +   52.00  974.26  1.4e-46   159  !
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-439-313-209 +   52.00  974.26  1.4e-46   159  !
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-37 +   51.00  955.36  1.6e-45   158  !
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-9 +   38.00  700.35  2.5e-31   489  !
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-10 +   37.00  681.14  3.0e-30   506  !
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-38 +   33.00  615.49  1.4e-26   136  !
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-5 +   24.00  432.00  2.2e-16   732  !
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-11 +   14.00  248.71  3.6e-06   329  !
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-12 +   14.00  248.71  3.6e-06   329  !
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-13 +   14.00  248.71  3.6e-06   329  !
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-14 +   14.00  248.71  3.6e-06   329  !
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-15 +   14.00  248.68  3.7e-06   330  !
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-42 +   11.00  207.38  0.0007    44  !
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-44 +   11.00  206.55  0.0008    49  !
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-40 -   11.00  192.30  0.0050   311  !
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-41 +   10.00  173.38  0.0572   310  !
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-09-023-082A-41 +    7.00  135.08  7.77      28  !
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-09-023-082A-4 +    7.00  133.36  9.68      35  !  S
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-09-023-082A-6 +    7.00  133.36  9.68      35  !  S
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-09-023-082A-12 +    7.00  133.36  9.68      35  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-757-653-121 -    7.00  117.30  75.99     281  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-757-653-127 +    7.00  117.30  75.99     281  !
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-171-461-51 +    7.00  116.20  87.48     324  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-756-386-56 -    7.00  115.90  90.95     337  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-823-516-45 -    7.00  115.90  90.95     337  !
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-682-853A-56 -    7.00  115.90  90.95     337  !
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-759-038-56 -    7.00  115.90  90.95     337  !
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-758-314-56 -    7.00  115.90  90.95     337  !
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-350-309-56 -    7.00  115.90  90.95     337  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-691-814B-113 -    7.00  115.48  96.02     356  !
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-020-956-160 -    7.00  114.97  102.42    380
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-030-607-160 -    7.00  114.97  102.42    380
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-439-313-160 -    7.00  114.97  102.42    380
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-794-002-5 -    7.00  112.35  143.38    534  !
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-854-039B-5 -    7.00  112.35  143.38    534
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-854-039B-1 -    7.00  111.50  159.84    596
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-406-248-3 -    7.00  111.49  160.10    597  !

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-1

seq_documentation_block:
; Sequence 1, Application US/08916576B
; Patent No. 6171816
;  GENERAL INFORMATION:
;    APPLICANT:  YU, GUO-LIANG
;    APPLICANT:  DILLON, PATRICK J.
;    APPLICANT:  EBNER, REINHARD
;    APPLICANT:  ENDRESS, GREGORY A.
;    TITLE OF INVENTION:  NOVEL HUMAN GROWTH FACTORS
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  US
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/916,576B
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/024,347
;      FILING DATE:  23-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  STEFFE, ERIC K.
;      REGISTRATION NUMBER:  36,688
;      REFERENCE/DOCKET NUMBER:  1488.0500001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  875 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  71..595
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  131..595
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  71..130
US-08-916-576B-1


alignment_scores:
            Quality:  192.00              Length:    192
              Ratio:  1.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x US-08-916-576B-1   ..
```

```
Align seg 1/1  to: US-08-916-576B-1  from: 1  to: 875

     15 ProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerArgValAl 31
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
     20 CCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAGAGTTGC 69

     31 aMetGluLysIleProValSerAlaPheLeuLeuLeuValAlaLeuSerT 48
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
     70 CATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCTCCT 119

     48 yrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLysAspThr 64
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    120 ACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGACACA 169

     65 LysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTrpGlyAs 81
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    170 AAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGGGTGA 219

     81 pGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerLysT 98
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    220 CCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAATCCAAGA 269

     98 hrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCysProHis 114
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    270 CAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGCCCACAC 319

    115 SerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGlnLysLe 131
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    320 AGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCAGAAATT 369

    131 uAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrAspLysH 148
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    370 GGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTGACAAAC 419

    148 isLeuSerProAspGlyGlnTyrValProArgIleMetPheValAspPro 164
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    420 ACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTTGACCCA 469

    165 SerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnArgLeuTy 181
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    470 TCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAATCGTCTCTA 519

    181 rAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMetLysLysA 198
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    520 TGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACATGAAGAAAG 569

    198 laLeuLysLeuLeuLysThrGluLeu 206
        |||||||||||||||||||||||||
    570 CTCTCAAGTTGCTGAAGACTGAATTG 595

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-247-155-61

seq_documentation_block:
; Sequence 61, Application US/09247155A
; Patent No. 6312922
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, Jean-Baptiste
;  APPLICANT: Duclert, Aymeric
;  APPLICANT: Bougueleret, Lydie
;  TITLE OF INVENTION: Complementary DNAs
;  FILE REFERENCE: GENSET.021A
;  CURRENT APPLICATION NUMBER: US/09/247,155A
;  CURRENT FILING DATE: 1999-02-09
;  EARLIER APPLICATION NUMBER: 60/074,121
;  EARLIER FILING DATE: 1998-02-09
;  EARLIER APPLICATION NUMBER: 60/081,563
;  EARLIER FILING DATE: 1998-04-13
;  EARLIER APPLICATION NUMBER: 60/096,116
;  EARLIER FILING DATE: 1998-08-10
;  EARLIER APPLICATION NUMBER: 60/099,273
;  EARLIER FILING DATE: 1998-10-04
;  NUMBER OF SEQ ID NOS: 182
;  SOFTWARE: Patent.pm

; SEQ ID NO 61
;   LENGTH: 1689
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 51..575
;   FEATURE:
;   NAME/KEY: sig_peptide
;   LOCATION: 51..110
;   OTHER INFORMATION: Von Heijne matrix
;   OTHER INFORMATION: score 11.2
;   OTHER INFORMATION: seq AFLLLVALSYTLA/RD
;   FEATURE:
;   NAME/KEY: polyA_signal
;   LOCATION: 1653..1658
;   FEATURE:
;   NAME/KEY: polyA_site
;   LOCATION: 1674..1689
US-09-247-155-61


alignment_scores:
           Quality:  178.00              Length:    178
             Ratio:  1.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x US-09-247-155-61    ..

 Align seg 1/1  to: US-09-247-155-61  from: 1  to: 1689

     29 ArgValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAl 45
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
     42 AGAGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGC 91

     45 aLeuSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysL 62
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
     92 CCTCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAA 141

     62 ysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGly 78
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    142 AGGACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGT 191

     79 TrpGlyAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLy 95
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    192 TGGGGTGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAA 241

     95 sSerLysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluC 112
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    242 ATCCAAGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGT 291

    112 ysProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGluIle 128
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    292 GCCCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATC 341

    129 GlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrTh 145
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    342 CAGAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAAC 391

    145 rAspLysHisLeuSerProAspGlyGlnTyrValProArgIleMetPheV 162
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    392 TGACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTG 441

    162 alAspProSerLeuThrValArgAlaAspIleThrGlyArgTyrSerAsn 178
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    442 TTGACCCATCTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAAT 491

    179 ArgLeuTyrAlaTyrGluProAlaAspThrAlaLeuLeuLeuAspAsnMe 195
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    492 CGTCTCTATGCTTACGAACCTGCAGATACAGCTCTGTTGCTTGACAACAT 541

    195 tLysLysAlaLeuLysLeuLeuLysThrGluLeu 206
```

```
          ||||||||||||||||||||||||||||||||||
      542 GAAGAAAGCTCTCAAGTTGCTGAAGACTGAATTG 575

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-35

seq_documentation_block:
; Sequence 35, Application US/08916576B
; Patent No. 6171816
;  GENERAL INFORMATION:
;    APPLICANT:  YU, GUO-LIANG
;    APPLICANT:  DILLON, PATRICK J.
;    APPLICANT:  EBNER, REINHARD
;    APPLICANT:  ENDRESS, GREGORY A.
;    TITLE OF INVENTION:  NOVEL HUMAN GROWTH FACTORS
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  US
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/916,576B
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/024,347
;      FILING DATE:  23-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  STEFFE, ERIC K.
;      REGISTRATION NUMBER:  36,688
;      REFERENCE/DOCKET NUMBER:  1488.0500001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  35:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  386 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-916-576B-35


alignment_scores:
            Quality:  107.00              Length:    107
              Ratio:  1.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x US-08-916-576B-35   ..

 Align seg 1/1  to: US-08-916-576B-35  from: 1  to: 386

      82 GlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerLysTh 98
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      64 CAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAATCCAAGAC 113

      98 rSerAsnLysProLeuMetIleIleHisHisLeuAspGluCysProHisS 115
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     114 AAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGCCCACACA 163

     115 erGlnAlaLeuLysLysValPheAlaGluAsnLysGluIleGlnLysLeu 131
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     164 GTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGAAATCCAGAAATTG 213

     132 AlaGluGlnPheValLeuLeuAsnLeuValTyrGluThrThrAspLysHi 148
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     214 GCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAACAACTGACAAACA 263

     148 sLeuSerProAspGlyGlnTyrValProArgIleMetPheValAspProS 165
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     264 CCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATGTTTGTTGACCCAT 313

     165 erLeuThrValArgAlaAspIleThrGlyArgTyrSerAsnArgLeuTyr 181
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     314 CTCTGACAGTTAGAGCCGATATCACTGGAAGATATTCAAATCGTCTCTAT 363

     182 AlaTyrGluProAlaAspThr 188
         ||||||||||||||||||||
     364 GCTTACGAACCTGCAGATACA 384

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-040-984-78

seq_documentation_block:
; Sequence 78, Application US/09040984
; Patent No. 6210883
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    APPLICANT:  Wang, TongTong
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR DIAGNOSIS
;    TITLE OF INVENTION:  OF LUNG CANCER
;    NUMBER OF SEQUENCES:  86
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/040,984
;      FILING DATE:  18-MAR-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.456
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-622-4900
;      TELEFAX:  206-282-6031
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  78:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  793 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-040-984-78


alignment_scores:
            Quality:  102.00              Length:    102
              Ratio:  1.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x US-09-040-984-78   ..

 Align seg 1/1  to: US-09-040-984-78  from: 1  to: 793

      14 HisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerArgVa 30
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       2 CATCCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAGAGT 51
```

```
     30 lAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAlaLeuS 47
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     52 TGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCT 101

     47 erTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLysAsp 63
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    102 CCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGAC 151

     64 ThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTrpGl 80
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    152 ACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGGG 201

     80 yAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerL 97
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    202 TGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAATCCA 251

     97 ysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCysPro 113
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    252 AGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGCCCA 301

    114 HisSer 115
        ||||||
    302 CACAGT 307

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-123-912-78

seq_documentation_block:
; Sequence 78, Application US/09123912A
; Patent No. 6312695
; GENERAL INFORMATION:
;  APPLICANT:  Reed, Steven G.
;  APPLICANT:  Wang, Tongtong
;  TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR THERAPY OF LUNG CANCER
;  FILE REFERENCE:  210121.455C1
;  CURRENT APPLICATION NUMBER: US/09/123,912A
;  CURRENT FILING DATE:  1998-07-27
;  PRIOR APPLICATION NUMBER:  09/040,802
;  PRIOR FILING DATE:  1998-03-18
;  NUMBER OF SEQ ID NOS:  114
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 78
;   LENGTH: 793
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (309)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (492)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (563)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (657)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (660)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (703)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (708)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (710)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (711)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (732)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (740)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (748)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (758)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (762)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (765)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (787)
;   OTHER INFORMATION: Where n is a, c, g or t
US-09-123-912-78


alignment_scores:
            Quality:  102.00            Length:    102
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x US-09-123-912-78   ..

 Align seg 1/1  to: US-09-123-912-78  from: 1  to: 793

     14 HisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerArgVa 30
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      2 CATCCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAGAGT 51

     30 lAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAlaLeuS 47
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     52 TGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCT 101

     47 erTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLysAsp 63
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    102 CCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGAC 151

     64 ThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGlyTrpGl 80
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    152 ACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGTTGGGG 201

     80 yAspGlnLeuIleTrpThrGlnThrTyrGluGluAlaLeuTyrLysSerL 97
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    202 TGACCAACTCATCTGGACTCAGACATATGAAGAAGCTCTATATAAATCCA 251

     97 ysThrSerAsnLysProLeuMetIleIleHisHisLeuAspGluCysPro 113
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    252 AGACAAGCAACAAACCCTTGATGATTATTCATCACTTGGATGAGTGCCCA 301

    114 HisSer 115
        ||||||
    302 CACAGT 307

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-221-298-32

seq_documentation_block:
; Sequence 32, Application US/09221298
; Patent No. 6284241
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY AND DIAGNOSIS
;  TITLE OF INVENTION: OF COLON CANCER
;  FILE REFERENCE: 210121.471
;  CURRENT APPLICATION NUMBER: US/09/221,298
```

```
;  CURRENT FILING DATE: 1998-12-23
;  NUMBER OF SEQ ID NOS: 112
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
;   LENGTH: 401
;   TYPE: DNA
;   ORGANISM: Human
US-09-221-298-32


alignment_scores:
            Quality:   66.00              Length:     80
              Ratio:  0.835                Gaps:      1
 Percent Similarity:  98.750   Percent Identity: 98.750

alignment_block:
 US-09-674-266A-181 x US-09-221-298-32   ..

 Align seq 1/1  to: US-09-221-298-32  from: 1  to: 401

     111 GluCysProHisSerGlnAlaLeuLysLysValPheAlaGluAsnLysGl 127
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     164 GAGTGCCCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAATAAAGA 213

     127 uIleGlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValTyrGluT 144
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     214 AATCCAGAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTTATGAAA 263

     144 hrThrAspLysHisLeuSerProAspGlyGlnTyrValProArgIleMet 160
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     264 CAACTGACAAACACCTTTCTCCTGATGGCCAGTATGTCCCCAGGATTATG 313

     161 PheValAspProSerLeuThrValArgAla.AspIleThrGlyArgTyrS 177
         |||||||||||||||||||||||||||||| ||||||||||||||||||
     314 TTTGTTGACCCATCTCTGACAGTTAGAGCCCGATATCACTGGAAGATATT 363

     177 erAsnArgLeuTyrAlaTyrGluProAlaAspThrAla 189
         |||||||||||||||||||||||||||||||||||||
     364 CAAACCGTCTCTATGCTTACGAACCTGCAGATACAGCT 401

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-247-155-149

seq_documentation_block:
; Sequence 149, Application US/09247155A
; Patent No. 6312922
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, Jean-Baptiste
;  APPLICANT: Duclert, Aymeric
;  APPLICANT: Bougueleret, Lydie
;  TITLE OF INVENTION: Complementary DNAs
;  FILE REFERENCE: GENSET.021A
;  CURRENT APPLICATION NUMBER: US/09/247,155A
;  CURRENT FILING DATE: 1999-02-09
;  EARLIER APPLICATION NUMBER: 60/074,121
;  EARLIER FILING DATE: 1998-02-09
;  EARLIER APPLICATION NUMBER: 60/081,563
;  EARLIER FILING DATE: 1998-04-13
;  EARLIER APPLICATION NUMBER: 60/096,116
;  EARLIER FILING DATE: 1998-08-10
;  EARLIER APPLICATION NUMBER: 60/099,273
;  EARLIER FILING DATE: 1998-10-04
;  NUMBER OF SEQ ID NOS: 182
;  SOFTWARE: Patent.pm
; SEQ ID NO 149
;   LENGTH: 940
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 177..569
;   FEATURE:
;   NAME/KEY: sig_peptide
;   LOCATION: 177..236
;   OTHER INFORMATION: Von Heijne matrix
;   OTHER INFORMATION: score 11.1999998092651
;   OTHER INFORMATION: seq AFLLLVALSYTLA/RD
;   FEATURE:
;   NAME/KEY: polyA_site
;   LOCATION: 931..939
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 482
;   OTHER INFORMATION: n=a, g, c or t
US-09-247-155-149


alignment_scores:
            Quality:   62.00              Length:     62
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x US-09-247-155-149   ..

 Align seq 1/1  to: US-09-247-155-149  from: 1  to: 940

      29 ArgValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAl 45
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     168 AGAGTTGCCATGGAAAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGC 217

      45 aLeuSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysL 62
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     218 CCTCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAA 267

      62 ysAspThrLysAspSerArgProLysLeuProGlnThrLeuSerArgGly 78
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     268 AGGACACAAAGGACTCTCGACCCAAACTGCCCCAGACCCTCTCCAGAGGT 317

      79 TrpGlyAspGlnLeuIleTrpThrGlnThrTyrGlu 90
         |||||||||||||||||||||||||||||||||||
     318 TGGGGTGACCAACTCATCTGGACACARACATATGAA 353

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-36

seq_documentation_block:
; Sequence 36, Application US/08916576B
; Patent No. 6171816
;  GENERAL INFORMATION:
;    APPLICANT:  YU, GUO-LIANG
;    APPLICANT:  DILLON, PATRICK J.
;    APPLICANT:  EBNER, REINHARD
;    APPLICANT:  ENDRESS, GREGORY A.
;    TITLE OF INVENTION:  NOVEL HUMAN GROWTH FACTORS
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  US
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/916,576B
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/024,347
;      FILING DATE:  23-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  STEFFE, ERIC K.
;      REGISTRATION NUMBER:  36,688
```

```
;      REFERENCE/DOCKET NUMBER:  1488.0500001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  36:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  373 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-916-576B-36


alignment_scores:
            Quality:   53.00             Length:     53
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x US-08-916-576B-36   ..

 Align seg 1/1  to: US-08-916-576B-36  from: 1  to: 373

      16 SerArgArgLeuThrGlnGlyArgTrpValArgLysSerArgValAlaMe 32
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 AGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAGAGTTGCCAT 50

      32 tGluLysIleProValSerAlaPheLeuLeuLeuValAlaLeuSerTyrT 49
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      51 GGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCTCCTACA 100

      49 hrLeuAlaArgAspThrThrValLysProGlyAlaLysLysAspThrLys 65
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 CTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGACACAAAG 150

      66 AspSerArg 68
         |||||||||
     151 GACTCTCGA 159

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-030-607-209

seq_documentation_block:
; Sequence 209, Application US/09030607
; Patent No. 6262245
;  GENERAL INFORMATION:
;    APPLICANT:  Xu, Jiangchun
;    APPLICANT:  Dillon, Davin C.
;    TITLE OF INVENTION:  COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO
;    NUMBER OF SEQUENCES:  224
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/030,607
;      FILING DATE:  25-FEB-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.427C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  209:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  159 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-030-607-209


alignment_scores:
            Quality:   52.00             Length:     52
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x US-09-030-607-209   ..

 Align seg 1/1  to: US-09-030-607-209  from: 1  to: 159

      25 ValArgLysSerArgValAlaMetGluLysIleProValSerAlaPheLe 41
         |||||||||||||||||||||||||||||||||||||||||||||||||||
       3 GTGAGGAAATCCAGAGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTT 52

      41 uLeuLeuValAlaLeuSerTyrThrLeuAlaArgAspThrThrValLysP 58
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      53 GCTCCTTGTGGCCCTCTCCTACACTCTGGCCAGAGATACCACAGTCAAAC 102

      58 roGlyAlaLysLysAspThrLysAspSerArgProLysLeuProGlnThr 74
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     103 CTGGAGCCAAAAAGGACACAAAGGACTCTCGACCCAAACTGCCCCAGACC 152

      75 LeuSer 76
         ||||||
     153 CTCTCC 158

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-439-313-209

seq_documentation_block:
; Sequence 209, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Mitcham, Jennifer L.
;  APPLICANT:  Harlocker, Susan Louise
;  APPLICANT:  Jiang Yuqui
;  APPLICANT:  Reed, Steven G.
;  APPLICANT:  Kalos, Michael
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Retter, Mark
;  APPLICANT:  Solk, John
;  APPLICANT:  Day, Craig
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF PROSTATE CANCER
;  FILE REFERENCE: 210121.427C9
;  CURRENT APPLICATION NUMBER: US/09/439,313
;  CURRENT FILING DATE:  1999-11-12
;  NUMBER OF SEQ ID NOS: 575
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 209
;   LENGTH: 159
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-439-313-209


alignment_scores:
            Quality:   52.00             Length:     52
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
```

```
 US-09-674-266A-181 x US-09-439-313-209   ..

 Align seg 1/1  to: US-09-439-313-209  from: 1  to: 159

      25 ValArgLysSerArgValAlaMetGluLysIleProValSerAlaPheLe 41
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       3 GTGAGGAAATCCAGAGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTT 52

      41 uLeuLeuValAlaLeuSerTyrThrLeuAlaArgAspThrThrValLysP 58
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      53 GCTCCTTGTGGCCCTCTCCTACACTCTGGCCAGAGATACCACAGTCAAAC 102

      58 roGlyAlaLysLysAspThrLysAspSerArgProLysLeuProGlnThr 74
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     103 CTGGAGCCAAAAAGGACACAAAGGACTCTCGACCCAAACTGCCCCAGACC 152

      75 LeuSer 76
         ||||||
     153 CTCTCC 158

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-37

seq_documentation_block:
; Sequence 37, Application US/08916576B
; Patent No. 6171816
;  GENERAL INFORMATION:
;    APPLICANT:  YU, GUO-LIANG
;    APPLICANT:  DILLON, PATRICK J.
;    APPLICANT:  EBNER, REINHARD
;    APPLICANT:  ENDRESS, GREGORY A.
;    TITLE OF INVENTION:  NOVEL HUMAN GROWTH FACTORS
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  US
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/916,576B
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/024,347
;      FILING DATE:  23-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  STEFFE, ERIC K.
;      REGISTRATION NUMBER:  36,688
;      REFERENCE/DOCKET NUMBER:  1488.0500001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  37:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  158 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-916-576B-37


alignment_scores:
            Quality:   51.00            Length:     51
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000
```

```
alignment_block:
 US-09-674-266A-181 x US-08-916-576B-37   ..

 Align seg 1/1  to: US-08-916-576B-37  from: 1  to: 158

      15 ProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerArgValAl 31
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       6 CCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAGAGTTGC 55

      31 aMetGluLysIleProValSerAlaPheLeuLeuLeuValAlaLeuSerT 48
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      56 CATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCCCTCTCCT 105

      48 yrThrLeuAlaArgAspThrThrValLysProGlyAlaLysLysAspThr 64
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     106 ACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAAAGGACACA 155

      65 Lys 65
         |||
     156 AAG 158

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-9

seq_documentation_block:
; Sequence 9, Application US/08916576B
; Patent No. 6171816
;  GENERAL INFORMATION:
;    APPLICANT:  YU, GUO-LIANG
;    APPLICANT:  DILLON, PATRICK J.
;    APPLICANT:  EBNER, REINHARD
;    APPLICANT:  ENDRESS, GREGORY A.
;    TITLE OF INVENTION:  NOVEL HUMAN GROWTH FACTORS
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  US
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/916,576B
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/024,347
;      FILING DATE:  23-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  STEFFE, ERIC K.
;      REGISTRATION NUMBER:  36,688
;      REFERENCE/DOCKET NUMBER:  1488.0500001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  489 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-916-576B-9


alignment_scores:
            Quality:   38.00            Length:     38
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000
```

```
alignment_block:
 US-09-674-266A-181 x US-08-916-576B-9   ..

 Align seg 1/1  to: US-08-916-576B-9  from: 1  to: 489

      29 ArgValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAl 45
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      40 AGAGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGC 89

      45 aLeuSerTyrThrLeuAlaArgAspThrThrValLysProGlyAlaLysL 62
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      90 CCTCTCCTACACTCTGGCCAGAGATACCACAGTCAAACCTGGAGCCAAAA 139

      62 ysAspThrLysAsp 66
         |||||||||||||
     140 AGGACACAAAGGAC 153

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-10

seq_documentation_block:
; Sequence 10, Application US/08916576B
; Patent No. 6171816
;  GENERAL INFORMATION:
;    APPLICANT:  YU, GUO-LIANG
;    APPLICANT:  DILLON, PATRICK J.
;    APPLICANT:  EBNER, REINHARD
;    APPLICANT:  ENDRESS, GREGORY A.
;    TITLE OF INVENTION:  NOVEL HUMAN GROWTH FACTORS
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  US
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/916,576B
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/024,347
;      FILING DATE:  23-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  STEFFE, ERIC K.
;      REGISTRATION NUMBER:  36,688
;      REFERENCE/DOCKET NUMBER:  1488.0500001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  506 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-916-576B-10


alignment_scores:
            Quality:   37.00             Length:     37
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x US-08-916-576B-10   ..

 Align seg 1/1  to: US-08-916-576B-10  from: 1  to: 506

     109 LeuAspGluCysProHisSerGlnAlaLeuLysLysValPheAlaGluAs 125
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      50 TTGGATGAGTGCCCACACAGTCAAGCTTTAAAGAAAGTGTTTGCTGAAAA 99

     125 nLysGluIleGlnLysLeuAlaGluGlnPheValLeuLeuAsnLeuValT 142
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     100 TAAAGAAATCCAGAAATTGGCAGAGCAGTTTGTCCTCCTCAATCTGGTTT 149

     142 yrGluThrThr 145
         ||||||||||
     150 ATGAAACAACT 160

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-38

seq_documentation_block:
; Sequence 38, Application US/08916576B
; Patent No. 6171816
;  GENERAL INFORMATION:
;    APPLICANT:  YU, GUO-LIANG
;    APPLICANT:  DILLON, PATRICK J.
;    APPLICANT:  EBNER, REINHARD
;    APPLICANT:  ENDRESS, GREGORY A.
;    TITLE OF INVENTION:  NOVEL HUMAN GROWTH FACTORS
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  US
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/916,576B
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/024,347
;      FILING DATE:  23-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  STEFFE, ERIC K.
;      REGISTRATION NUMBER:  36,688
;      REFERENCE/DOCKET NUMBER:  1488.0500001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  38:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  136 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-916-576B-38


alignment_scores:
            Quality:   33.00             Length:     33
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x US-08-916-576B-38   ..

 Align seg 1/1  to: US-08-916-576B-38  from: 1  to: 136
```

```
     13 ProHisProSerArgArgLeuThrGlnGlyArgTrpValArgLysSerAr 29
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      2 CCGCATCCTAGCCGCCGACTCACACAAGGCAGGTGGGTGAGGAAATCCAG 51

     29 gValAlaMetGluLysIleProValSerAlaPheLeuLeuLeuValAla 45
        |||||||||||||||||||||||||||||||||||||||||||||||||
     52 AGTTGCCATGGAGAAAATTCCAGTGTCAGCATTCTTGCTCCTTGTGGCC 100

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-916-576B-5

seq_documentation_block:
; Sequence 5, Application US/08916576B
; Patent No. 6171816
;  GENERAL INFORMATION:
;    APPLICANT:  YU, GUO-LIANG
;    APPLICANT:  DILLON, PATRICK J.
;    APPLICANT:  EBNER, REINHARD
;    APPLICANT:  ENDRESS, GREGORY A.
;    TITLE OF INVENTION:  NOVEL HUMAN GROWTH FACTORS
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  US
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/916,576B
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/024,347
;      FILING DATE:  23-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  STEFFE, ERIC K.
;      REGISTRATION NUMBER:  36,688
;      REFERENCE/DOCKET NUMBER:  1488.0500001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  732 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  49..546
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  118..546
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  49..117
US-08-916-576B-5


alignment_scores:
            Quality:   24.00              Length:     24
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-674-266A-181 x US-08-916-576B-5    ..
```

```
Align seg 1/1  to: US-08-916-576B-5  from: 1  to: 732

    149 LeuSerProAspGlyGlnTyrValProArgIleMetPheValAspProSe 165
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    373 TTATCACCTGATGGGCAATATGTGCCTAGAATCATGTTTGTAGACCCTTC 422

    165 rLeuThrValArgAlaAspIle 172
        |||||||||||||||||||||
    423 TTTAACAGTTAGAGCTGACATA 444
```